

STATUS OF MACHINE TRANSLATION (MT) TECHNOLOGY

HEARING BEFORE THE SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY OF THE COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY U.S. HOUSE OF REPRESENTATIVES ONE HUNDRED FIRST CONGRESS SECOND SESSION

SEPTEMBER 11, 1990

[No. 153]

Printed for the use of the
Committee on Science, Space, and Technology



U.S. GOVERNMENT PRINTING OFFICE

WASHINGTON : 1990

35-075

For sale by the Superintendent of Documents, Congressional Sales Office
U.S. Government Printing Office, Washington, DC 20402

H701-14

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

ROBERT A. ROE, *New Jersey, Chairman*

GEORGE E. BROWN, Jr., California
JAMES H. SCHEUER, New York
MARILYN LLOYD, Tennessee
DOUG WALGREN, Pennsylvania
DAN GLICKMAN, Kansas
HAROLD L. VOLKMER, Missouri
HOWARD WOLPE, Michigan
BILL NELSON, Florida
RALPH M. HALL, Texas
DAVE McCURDY, Oklahoma
NORMAN Y. MINETA, California
TIM VALENTINE, North Carolina
ROBERT G. TORRICELLI, New Jersey
RICK BOUCHER, Virginia
TERRY L. BRUCE, Illinois
RICHARD H. STALLINGS, Idaho
JAMES A. TRAFICANT, Jr., Ohio
LEE H. HAMILTON, Indiana
HENRY J. NOWAK, New York
CARL C. PERKINS, Kentucky
TOM McMILLEN, Maryland
DAVID E. PRICE, North Carolina
DAVID R. NAGLE, Iowa
JIMMY HAYES, Louisiana
DAVID E. SKAGGS, Colorado
JERRY F. COSTELLO, Illinois
HARRY JOHNSTON, Florida
JOHN TANNER, Tennessee
GLEN BROWDER, Alabama

ROBERT S. WALKER, Pennsylvania*
F. JAMES SENSENBRENNER, Jr.,
Wisconsin
CLAUDINE SCHNEIDER, Rhode Island
SHERWOOD L. BOEHLERT, New York
TOM LEWIS, Florida
DON RITTER, Pennsylvania
SID MORRISON, Washington
RON PACKARD, California
PAUL B. HENRY, Michigan
HARRIS W. FAWELL, Illinois
D. FRENCH SLAUGHTER, Jr., Virginia
LAMAR SMITH, Texas
JACK BUECHNER, Missouri
CONSTANCE A. MORELLA, Maryland
CHRISTOPHER SHAYS, Connecticut
DANA ROHRBACHER, California
STEVEN H. SCHIFF, New Mexico
TOM CAMPBELL, California
JOHN J. RHODES, III, Arizona

ROBERT C. KETCHAM, *Chief of Staff*
JOHN S. DOYLE, JR., *Chief Counsel*
CAROLYN C. GREENFELD, *Chief Clerk*
DAVID D. CLEMENT, *Republican Chief of Staff*

SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY

TIM VALENTINE, *North Carolina, Chairman*

GEORGE E. BROWN, Jr., California
HOWARD WOLPE, Michigan
LEE H. HAMILTON, Indiana
DAVID E. PRICE, North Carolina
NORMAN Y. MINETA, California
TERRY L. BRUCE, Illinois
DAVID R. NAGLE, Iowa
DAVID E. SKAGGS, Colorado
JERRY F. COSTELLO, Illinois
HARRY JOHNSTON, Florida
JIMMY HAYES, Louisiana
GLEN BROWDER, Alabama

SHERWOOD L. BOEHLERT, New York
CLAUDINE SCHNEIDER, Rhode Island
DON RITTER, Pennsylvania
PAUL B. HENRY, Michigan
CONSTANCE A. MORELLA, Maryland
TOM CAMPBELL, California
D. FRENCH SLAUGHTER, Jr., Virginia
JACK BUECHNER, Missouri
JOHN J. RHODES, III, Arizona

*Ranking Republican Member.

CONTENTS

WITNESSES

September 11, 1990:

Martha Caldwell Harris, Director, Office of Japan Affairs, National Research Council, Washington, DC, and Muriel Vasconcellos, Chief of Translation, Pan American Health Organization, Washington, DC	11
Deborah L. Wince-Smith, Assistant Secretary for Technology Policy, Department of Commerce, Washington, DC, Accompanied by Joseph E. Clark, Co-chair of Task Force on Access to Scientific and Technical Information, and Deputy Director, National Technical Information Service, Department of Commerce; Charles Brownstein, Acting Assistant Director, Computer and Information Science and Engineering, NSF, Washington, DC, accompanied by Y.T. Chien, Division Director, Information, Robotics, and Intelligent Systems; and Dale Bostad, Technical Advisor, Directorate of Translation, Foreign Technology Division, U.S. Air Force, Wright Patterson Air Force Base, Ohio.....	53
Jaime Carbonell, Director, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, Pennsylvania; Winfield S. Bennett, Linguistics Research Center, The University of Texas at Austin; David E. Johnson, Research Scientist, IBM Corporation, Thomas J. Watson Research Laboratories, and Gregory Zaretsky, President, The Corporate Word, Inc., Pittsburgh, Pennsylvania.....	115

APPENDIX

Report of a National Research Council Symposium on "Japanese-to-English Machine Translation" held December, 1989	167
U.S.-Japan Task Force Report on Scientific and Technical Information, April 1990	214
Department of Commerce Report to Congress, July 1988, "Current Status of Japanese-to-English Machine Translation"	237
Article Published in <i>Insight Magazine</i>	254
Article in <i>Time Magazine</i> July 24, 1989, "Trying to Decipher Babel"	263

STATUS OF MACHINE TRANSLATION (MT) TECHNOLOGY

TUESDAY, SEPTEMBER 11, 1990

HOUSE OF REPRESENTATIVES,
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY,
SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY,
Washington, D.C.

The subcommittee met, pursuant to call, at 9:35 a.m., in Room 2318, Rayburn House Office Building, Hon. Tim Valentine [chairman of the subcommittee] presiding.

Mr. VALENTINE. We will get started, folks, if you will please take your seats.

Today the Subcommittee on Science, Research and Technology will review the current status of technology to assist in the translation of materials from one language to another. Our witnesses will also discuss the future prospects of this technology and assess what, if any, the Federal Government's role should be in furthering translation research.

The technical community defines "machine translation"—that is (MT) as translation generated by a computer. Machine-aided translation (MAT) refers to human-initiated translation in which the computer is enlisted as an aid, mainly in the form of word processing, grammar checks, and terminology banks. Currently there are machine-aided translations on the market, but pure machine translation remains a goal.

Why is machine translation technology important?

In recent years, the need to keep abreast of scientific and technological developments abroad, together with an increasingly competitive global market, has generated a need for quick, accurate, and cost-effective translations.

The worldwide information explosion, a lack of qualified translators, plus the costliness associated with human translation efforts has heightened the importance of translation technology. The Japanese Government and several major Japanese companies have made the development of machine translation technology a priority.

In contrast to the Japanese, U.S. research in machine technology has been inactive. Numerous translation research projects were undertaken at United States universities in the 1950s and 1960s. They were encouraged and often financially supported by the United States Government, especially the Defense Department, which needed to expedite the translation of Russian technical material. However, the publication of a report by the Automatic Language

Processing Advisory Committee—that is ALPAC—of the National Academy of Sciences' National Research Council, in 1966, recommending the termination of governmental support for the machine translator projects, all but ended U.S. research in this field.

Interest and activity in machine translation technology have slowly revived. This morning we hope to gain some insight from our expert witnesses as to its present viability. Last December, the Commerce Department's Office of Japanese Technical Literature sponsored a national conference at the National Academy of Sciences on Japanese-to-English machine translation. A brief summary of those proceedings will be presented this morning.

We will also hear about a special study that's being launched by the U.S. Task Force on Access to Scientific and Technical Information to evaluate the status of Japanese research and development in machine translation technology.

We are fortunate to have with us today a number of leading experts in the United States in machine and machine-aided translation technologies. We look forward to exploring with our witnesses the promise of these technologies and whether the Federal Government should be doing more to promote research in this area.

I recognize at this time the distinguished former chairman of this subcommittee, Mr. Doug Walgren.

Mr. WALGREN. Thank you, Mr. Chairman.

I especially want to commend you and underscore the importance of this area in this committee's jurisdiction and express my appreciation for the attention that you have given to it. It's clear, I think, to all of us, that the Japanese-United States relationship is certainly second in importance to none in the present world, and will only grow more so as we go into the 1990s.

This relationship is a positive one, but nonetheless, one that is fraught with many difficulties. When I go to my town meetings in my district, I am presented with veterans of World War II who are quick to show their frustration, or vent their frustration at the fact that the grandchildren of the enemy that the United States defeated in World War II, and which they risked their lives in combat, that the grandchildren of their enemy have prospects that seem even better than the prospects of their own grandchildren, even though we have aided them so completely; or, and in fact we have aided them so completely in their recovery. And that veneer lies behind this relationship and makes it so important that the United States and Japan relationship succeed.

It is not succeeding now. When we have a trade deficit of 50-some billion dollars a year, it is a lightning rod for being blamed for unemployment in many of our industries; and when the Japanese are now investing in their factories, in their capital equipment more than we are—even though we are several times larger in population—when the benefits of that investment start to accrue to the wealth and productivity of Japan, and we have not made similar investments in our country, that relationship may be brought into—the questions about that relationship may rise even more sharply.

I say that to underscore the importance that what we do in this relationship is very, very essential for it to work out properly. And the ability to communicate, particularly with respect to technical

literature, is at the heart of being able to carry off this relationship successfully rather than to have it backslide into adversity feelings.

And we are starting at somewhat of a disadvantage because in our country we do not have the language skills that have been developed in Japan over the last several decades. In Japan it is hard to find a technical person who does not speak English and who does not have a history of spending time in United States universities.

In the United States it is hard to find a technical person who does speak Japanese and who has any contact whatsoever with Japan in terms of spending time there with that culture. So we have a lot of work to do on our side on this relationship.

Mr. Chairman, I would like to enter into the record three letters which I've sent to the forerunner, the National Institutes of Standards and Technology, urging them to carry out first an assessment and then the development of a plan to pursue machine translation in view of the importance of this relationship. The first dated April 6th, 1988; the second dated February 13th, 1989, directed to the Secretary of Commerce, Mr. Mosbacher; and again on August 22nd of this year, again directed to the Secretary of Commerce, Mr. Mosbacher. And I—stressing and asking in each instance that we first make a complete assessment and; secondly, that we develop a plan that we could pursue that would develop this area for us.

I want to underscore that we haven't made much progress in that. The Administration's witness will come forward today and say essentially that the next step is to assess the state of the art in Japan, which tells me that almost three years after our first letters we haven't gotten the assessment stage completed, let alone the planning stage.

And I know that this is a long-term project, but there—the efforts in the past Administration to emphasize Japanese Technical Literature and to increase the contact and the outreach on our side of this equation have not been willingly taken. And I hope that this Committee can play some role in keeping this new Administration focused on the potential in this area and the importance in this area.

And I would like to ask that those letters be included as part of my statement for the record and—

Mr. VALENTINE. Without objection.

Mr. WALGREN. I look forward to the testimony of the witnesses. Thank you, Mr. Chairman.

[The letters follow:]

DOUG WALGREN
18th District, Pennsylvania

COMMITTEES
ENERGY AND COMMERCE
SCIENCE AND TECHNOLOGY

Congress of the United States
House of Representatives
Washington, DC 20515

WASHINGTON OFFICE
2241 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515
(202) 225-2138

DISTRICT OFFICE
2117 FEDERAL BUILDING
1000 LIBERTY AVENUE
PITTSBURGH, PA 15222
(412) 381-4016

April 6, 1988

NIT

Mr. Ray Kammer
Deputy Director
National Bureau of Standards
Gaithersburg, Maryland 20899

Dear Mr. Kammer: *Ray*

This is a follow-up to the brief discussion we had concerning machine translation at the NBS hearing held on March 15, 1988

I was very pleased to hear that the Bureau is taking initial steps to conduct a national assessment of this area. Common sense certainly tells us that computer based translation has a tremendous contribution to make in keeping the U.S. current in science and technology. While we should continue encouraging the mastery of foreign languages by American engineers and scientists, in my view, we should as a nation give serious emphasis to developing a commercially viable machine translation system.

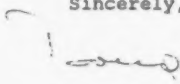
(2) I would urge the National Bureau of Standards to take the lead in organizing an assessment of the state-of-the-art in this area and creating a program which would stimulate both the public and private sector in its pursuit. As you know, a machine translation summit conference was held in Tokyo in the fall of 1987. I understand that another such conference will probably be held in Germany perhaps in fall 1989. I wonder if it would not make sense for the U.S. to sponsor a conference focused on Japanese-English translation possibility in the spring of 1989 in Washington, approximately one year from now.

I would hope a sound assessment process would result in a plan with a reasonable budget estimate showing how the various groups might contribute to this effort and that some provision could be made for these efforts in the FY 1990 budget cycle.

Page Two
April 6, 1988
Mr. Ray Kammer

I appreciate your interest in the contribution that could be made to machine translation development by NBS.

Sincerely,



Doug Walgren
Member of Congress

DW:fe

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
U.S. HOUSE OF REPRESENTATIVES
WASHINGTON, DC 20515

February 13, 1989

Honorable Robert Mosbacher
Secretary
Department of Commerce
Washington, DC

Dear Mr. Secretary:

As the new Administration prepares to set its priorities, I would like to underscore the urgent need to pursue machine translation capability to facilitate American access to Japanese Technical Literature.

In recognition of the need to improve access to Japan's scientific and technological information, the Committee on Science, Space, and Technology was instrumental in the passage of the Japanese Technical Literature Act of 1986 (P.L. 99-382). This legislation requires the Department of Commerce, through the National Technical Information Service, to expand and strengthen its efforts in gathering Japanese scientific and technological information.

As you know, one of the biggest obstacles in this effort is the language barrier. The Japanese language is recognized as being one of the most difficult for Westerners to master. While we should continue encouraging the mastery of foreign languages by American engineers and scientists, clearly the valuable information contained in Japanese scientific and technical literature will not be widely available in the U.S. until a commercially viable machine translation system has been developed.

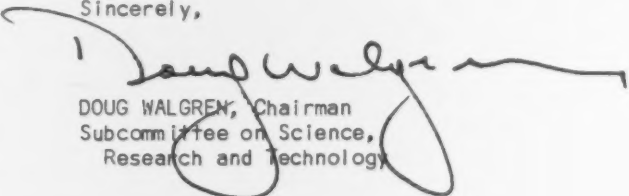
Last April, I addressed a letter to the National Institute of Standards and Technology (then the National Bureau of Standards) calling for an assessment of the present state of machine aided translation in the U.S. and the formulation of a national plan to hasten development of this crucial technology. In September, the Office of Japanese Technical Literature held an interagency meeting to discuss the status of machine translation.

A subsequent report of this interagency meeting concludes with the recommendation that "the Office of Japanese Technical Literature in cooperation with NTIS and other relevant government agencies and the private sector should develop a national plan for developing a Japanese-English Machine Translation System that is accessible to U.S. engineers and scientists."

Honorable Robert Mesbacher
February 13, 1988
Page 2

As you know, the National Institute of Standards and Technology Authorization Act for FY 1989 (P.L. 100-519) places the Japanese Technical Literature Office under the National Technical Information Service. This change reflects the intent of the authorizing Committees of the House and Senate to give this program a high profile in the new Technology Administration. I hope you will take a personal interest in moving our machine translation efforts forward under your administration. I would look forward to any opportunities to be helpful in this regard.

Sincerely,



DOUG WALGREN, Chairman
Subcommittee on Science,
Research and Technology

DW/Peg

DOUG WALGREN
19th District, Pennsylvania

COMMITTEE
ENERGY AND COMMERCE
SCIENCE AND TECHNOLOGY

Congress of the United States
House of Representatives
Washington, DC 20515

August 22, 1990

WASHINGTON OFFICE
2241 RAYMOND HOUSE OFFICE BUILDING
WASHINGTON, DC 20019
(202) 226-2126

DISTRICT OFFICE
2117 PENNSYLVANIA BUILDING
1000 LIBERTY AVENUE
PITTSBURGH, PA 15222
(412) 391-6018

The Honorable Robert Mosbacher
Secretary
U.S. Department of Commerce
14th & Constitution NW
Washington, D.C. 20230

Dear Mr. Secretary:

On September 11, 1990, the Subcommittee on Science, Research and Technology will hold hearings on machine translation (MT) as one means of accelerating the understanding and, most importantly, American access and use of Japanese science and technology.

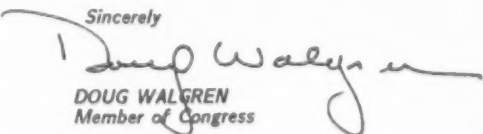
I have been greatly concerned with the need for the United States to develop capability for machine translation from Japanese to English. As we know, the Japanese, for their own purposes, have made a substantial effort to create a practical method of machine translation. Depending on how we assess their efforts, Japan has made major strides toward creating viable systems in this area.

Because of my long standing interest in this matter, I sent a letter on April 16, 1988 to the National Institute of Standards and Technology (then the National Bureau of Standards) requesting (1) a national assessment of machine translation efforts in the U.S. (including a comparison with Japanese and European efforts and, (2) a plan (including a budget) which would lay out the various steps that would be necessary to develop viable MT systems in the U.S. This request was based on the assumption that there would be close cooperation among government, industry and universities in making this assessment, and that any future effort would be a joint effort of all parties concerned in view of the importance of this tool for American competitiveness.

In response to my 1988 letter, I understand several interagency meetings have been held, along with a special symposium on machine translation convened in December 1989 by the National Academy of Sciences at the request of the Department of Commerce and that a report has been issued on this Symposium. Although, indirect, these responses are constructive and have developed a foundation that should be the basis for concrete steps in our development of machine translation.

In view of the importance of our relationship with Japan, I feel it critical to renew my request for (1) a national assessment of machine translation, analyzing where we, as a country stand, in developing MT systems and, (2) a specific plan (including preliminary budget estimates) detailing the steps necessary to reach this goal. I look forward to discussing this area with representatives of the Administration at the hearing on September 11, 1990.

Sincerely


DOUG WALGREN
Member of Congress

Mr. VALENTINE. I recognize the distinguished colleague from Rhode Island, Congresswoman Claudine Schneider.

Ms. SCHNEIDER. I appreciate your recognizing me, Mr. Chairman. But more importantly, I appreciate your convening this particular hearing this morning. And needless to say, we have much to learn.

I have no elaborate opening comments, and I'm looking forward to hearing from the witnesses.

Mr. VALENTINE. Thank you very much.

[The prepared opening statement of Mr. Costello follows:]

OPENING STATEMENT

BY U.S. REP. JERRY F. COSTELLO, ILLINOIS

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY

HEARING ON STATUS OF MACHINE TRANSLATION TECHNOLOGY

SEPTEMBER 11, 1990

Mr. Chairman, this hearing will focus on whether or not the federal involvement with machine translation and machine-aided translation should be expanded.

The United States is facing an extremely competitive international marketplace, and science technology is one area in which we have lost ground. Japan and European communities continue to invest millions into the improvement of their own high-performance industry. As worldwide technology continues to grow, so does the need for translated machines.

Mr. Chairman, I look forward to finding out whether the Federal Government should be doing more to promote research on machine and machine-aided translation. I thank you for your work in this area.

Mr. VALENTINE. I want to welcome all of the witnesses here—we have three panels—and those who also came to watch and listen.

The first panel consists of Dr. Martha Caldwell Harris, Director, Office of Japan Affairs, National Research Council, Washington, and Dr. Muriel Vasconcellos, Chief of Translation, Pan American Health Organization, in Washington.

Welcome. We—your prepared statements will appear in the record as presented to us, and I would ask you to—if you will please summarize or otherwise proceed as you deem appropriate. Welcome.

STATEMENTS OF DR. MARTHA CALDWELL HARRIS, DIRECTOR, OFFICE OF JAPAN AFFAIRS, NATIONAL RESEARCH COUNCIL, WASHINGTON, D.C., AND DR. MURIEL VASCONCELLOS, CHIEF OF TRANSLATION, PAN AMERICAN HEALTH ORGANIZATION, WASHINGTON, D.C.

Dr. HARRIS. Thank you, Mr. Chairman, for this opportunity to present testimony this morning. Over the years, this subcommittee has taken—

Mr. VALENTINE. Would you—let me see if that microphone is on, or if you would pull it a little bit closer. I don't think it's on.

Dr. HARRIS. Is it on now?

Mr. VALENTINE. It is now, yes.

Dr. HARRIS. Thank you. I am Martha Harris and I appreciate this opportunity to present testimony this morning.

Over the years, this Subcommittee has taken special interest in improving access to Japanese scientific and technical information.

Today's hearing is a good opportunity to look in some detail at an area of R&D—machine translation and related technologies—that may help to transcend the asymmetries in access to technical information between the U.S. and Japan. Despite significant efforts in both Japan and the United States over the past five years, the fundamental problem persists; the sheer numbers of Japanese students and researchers in the United States dwarf the comparable numbers of Americans in Japan and only a minute percentage of our technical personnel speak or read Japanese.

In view of Japan's status as a technological superpower, it is appropriate that policymakers concerned about our competitiveness and our security find new ways to access developments in Japan.

Machine translation is not a panacea or a fully developed technology, but it may provide tools for addressing some of these issues.

First, let me explain the role of the National Research Council in organizing a symposium on Japanese to English machine translation last December, and Japanese to English MT will be the focus of my remarks.

As a private nonprofit organization, we respond to requests from the U.S. Government to conduct studies, seminars and programs on science and technology.

Last year the Department of Commerce asked me, as the Director of the Office of Japan Affairs, to organize a symposium to bring together individuals from the private sector to examine recent developments in Japanese to English MT. We enlisted the coopera-

tion of the Computer Science and Technology Board of the National Research Council.

And, Mr. Chairman, I am pleased that this morning Marjorie Blumenthal, the staff director, for the Board, joins me this morning.

Working together with the Board in organizing the symposium, our principal goal was to provide a forum where individuals from the private sector, particularly potential developers and users in industry, could hear from researchers and others with firsthand experience. We were pleased that the response to the symposium was so positive, and that participants included individuals from a broad range of organizations. And a member of your Committee, Congressman Brown, addressed the symposium.

Mr. Chairman, a report of the symposium has been prepared and copies are available here today. Let me, therefore, touch on just a few highlights from the symposium and our related work.

First, while the early work on machine translation began in the United States, and excellent research continues here today, Japan has taken a lead in efforts to apply and commercialize Japanese to English machine translation technologies.

As one of the participants in our symposium put it, the United States produces the research papers while Japan produces the products. In contrast to the situation in the U.S., every major Japanese computer or electronics firm has invested considerable effort in machine translation R&D and many advertise commercial products.

As you noted earlier, the Japanese government is supporting machine translation R&D. Meanwhile, development of Japanese to English machine translation in the United States has lagged.

A second point: The needs of potential users of Japanese to English machine translation in the U.S. are in some ways different from those of the typical Japanese user. The typical user in Japan is a large corporation engaged in global operations. The Japanese user can control the source document because it is in his or her native language, can more easily tolerate marginal machine translation output because the company provides direct cost justification, and because of the more narrow domain of usage.

The potential user in the United States often has no fluency in Japanese, little control over source texts, and often needs more accurate and reliable outputs from broad domains rather than narrow technical fields.

In the opinion of some U.S. experts, the intermediate machine translation systems developed in Japan do not fit precisely the needs of many users in the U.S. for high quality, broad domain systems that may still be 10 to 20 years over the horizon.

A third point: Unfortunately, there is no generally accepted method for evaluating the quality and accuracy of machine translation, despite a vigorous debate about the effectiveness of systems on the market. Overall assessments of machine translation are less useful than evaluations of specific systems by particular users, because evaluation depends very much on the needs of the specific user. And experience shows that the more a user uses a system, the more efficient the work becomes.

I should say, though, that a major obstacle to machine translation development is the reluctance of developers to provide detailed information about performance characteristics and to exchange information about their experience. Therefore, potential users are well advised to conduct systematic comparisons on texts of their own choosing.

A fourth point: While more than 30 years of development work in a variety of language pairs has not produced a machine that replaces the human translator; today machine translation and related technologies are increasingly seen to augment the work of human translators.

Most of the Japanese to English machine translation systems now in use in Japan succeed because they are limited to particular domains, such as technical manuals for automobile companies. In addition, or as you said, a variety of machine aids for human translators that have practical uses today even if the prospect of a fully automated, high quality, low cost, general purpose system is over the horizon.

However, as mentioned above, the Japanese to English machine translation systems under development in Japan do not meet all the potential needs of U.S. users.

The question then is what should be done and who should do it?

The participants in our symposium shared a range of views on this subject. Skeptics wonder whether there is real demand for Japanese technical information and they, quite frankly, question whether the government should place priority here.

But a contrasting view is based on the recognition that markets sometimes fail to meet the national interest; in fact, the Japanese approach has been to consciously create demand.

As a technical challenge, Japanese to English machine translation represents an engineering stretch that may spin off progress in other areas of information technology. It is a test bed for solving natural language processing problems.

In addition, a focused effort bringing together researchers, developers, and users could provide an experiment for new approaches to collaborative R&D.

So if policymakers in the U.S. decide to support the development of machine translation technologies, there are a variety of approaches that could be taken. First, international cooperation in machine translation could focus on the precompetitive nature of many research themes. This could also speed dictionary development, text sharing and database construction.

An international consortium has been proposed and machine translation has been identified as an area for cooperation under the U.S./Japan Agreement for Cooperation in Science and Technology. U.S.-Japan cooperation could be particularly meaningful if it led to expanded cooperation with some of Japan's premier corporate programs.

Second, the U.S. Government could take the initiative in developing a U.S. strategy. The basic rationale for such this approach might be that there are no organizations in the U.S. with the resources to do this independently.

One participant in our symposium proposed a three-stage program that would involve the development of prototype Japanese to

English systems by competing consortia involving government, university and industry researchers.

In stage one a steering committee could identify bottlenecks for development, evaluate criteria for assessing the quality of the systems, and develop a plan for an organizational approach.

Stage two could focus on the development of two prototype systems. And in the final stage, interim products could be developed by subcontractors who focus efforts on machine translation for a wide variety of specific domains.

A purely private sector consortium is another possibility, but I should mention that a number of participants in the symposium said that the government would need to play some kind of role as catalyst.

Third, initiatives can also be taken by companies, corporations and other organizations. In order to build a constituency of users for Japanese technical information, the organization must invest the resources necessary in working with the users to tailor search, selection and output to their needs. But today, unfortunately, only the largest U.S. organizations have developed systematic approaches.

These kinds of efforts that I've outlined are not mutually exclusive. But it is not clear that international cooperation alone would build a sound U.S. research and production base for a new generation of Japanese to English systems.

If policymakers decide to pursue new initiatives, it will be important to consider ways to reduce the fragmentation that today characterizes to a great extent relationships among the research, development and user communities in our country.

Thank you.

[The prepared statement of Dr. Harris follows:]

TESTIMONY BY DR. MARINA CALDWELL HARRIS

DIRECTOR, OFFICE OF JAPAN AFFAIRS

NATIONAL RESEARCH COUNCIL

PREPARED FOR THE

SCIENCE, RESEARCH AND TECHNOLOGY SUBCOMMITTEE

HOUSE SCIENCE, SPACE AND TECHNOLOGY COMMITTEE

HEARING ON THE STATUS OF MACHINE TRANSLATION TECHNOLOGY,

SEPTEMBER 11, 1990

Thank you, Mr. Chairman for this opportunity to present testimony on "The Status of Machine Translation Technology." Over the years this subcommittee has taken special interest in improving access to Japanese scientific and technical information. Today's hearing is a good opportunity to look in some detail at an area of R&D— machine translation and related technologies— that may help to transcend the asymmetries in access to technical information between the United States and Japan. As you know, Mr. Chairman, despite significant efforts in both Japan and the United States over the past five years, the fundamental problem persists; the sheer numbers of Japanese students and researchers in the United States dwarf the comparable numbers of Americans in Japan and only a minute percentage of our technical personnel speak or read Japanese. In view of Japan's status as a technological superpower, it is appropriate that policymakers concerned about our competitiveness and security find new ways to improve access to developments in Japanese science and technology, including technology transfers to the United States through various mechanisms. Machine translation is not a panacea or a fully developed technology-- but it may provide tools for addressing some of these issues. My comments will focus on the status of Japanese to English machine translation, and whether there is a need for a U.S. effort in this area.

First, let me explain the role of the National Research Council in organizing a symposium on Japanese to English Machine Translation. The National Research Council is the operating arm of the National Academies of Sciences and Engineering and the Institute of Medicine. As a private non-

profit organization, we respond to requests from the US Government to conduct studies, seminars and programs concerning science and technology. I should mention that a new Office of Japan Affairs was established within the NRC in 1988 to address issues in US-Japan science and technology relations and to promote better working relationships between technical personnel in the two countries. During the past two years the Office of Japan Affairs has worked with the National Research Council's Committee on Japan, a group of distinguished experts from the private sector, to organize a variety of meetings and activities focusing on issues such as asymmetries in access and differences in the R&D systems.

Last year the Department of Commerce asked me as Director of the Office of Japan Affairs to organize a symposium to bring together individuals from the private sector to examine recent developments in Japanese to English Machine Translation. We enlisted the cooperation of the Computer Science and Technology Board of the NRC in this endeavor. The Computer Science and Technology Board (now the Computer and Telecommunications Board) was established in 1986 to address technical and policy issues pertaining to computer science, telecommunications, and associated technologies. It has initiated and responded to requests from the government for projects addressing the health of computing and communications fields and the associated industry base.

Working together with the CSTB in organizing the symposium, our principal goal was to provide a forum where individuals from the private sector, particularly potential developers and users in industry, could hear

from researchers and others with first hand experience. We were pleased that the response to the symposium was so positive, and that participants included individuals from a broad range of organizations.

Mr. Chairman, a report of the symposium has been prepared and copies are available here today. Let me touch on a few highlights from the symposium and our related research and follow-up activities.

1. While the early work on machine translation began in the United States and excellent research continues, Japan has taken a lead in efforts to apply and commercialize Japanese to English machine translation technologies. As one of the participants in our symposium put it, the United States produces research papers while Japan produces the products. In contrast to the situation in the United States, every major Japanese computer or electronics firm has invested considerable effort in machine translation R&D and many advertise commercial products. The Japanese government is supporting machine translation R&D. Meanwhile, development of Japanese to English machine translation in the United States has lagged. US researchers at the forefront of theoretical work may find it necessary to learn from those (in Japan) who take a more applications-oriented approach if the commercialization of useful systems is the goal.

2. The needs of potential users of Japanese to English machine translation in the United States are in some ways different from those of the typical Japanese user. The typical user in Japan is a large corporation engaged in global operations. The Japanese user can control the source

document because it is in his or her native language, can more easily tolerate marginal machine translation output because the company provides direct cost justification, and because of the narrow domain of usage. The potential user in the United States has no fluency in Japanese, no control over source texts, and needs more accurate and reliable output from broad domains. In the opinion of some US experts, the intermediate machine translation systems developed in Japan do not fit the needs of many US users for high quality, broad domain systems that are still 10 to 20 years over the horizon.

3. Unfortunately, there is no generally accepted method for evaluating the quality and accuracy of machine translation, despite a vigorous debate about the effectiveness of systems on the market. Overall assessments of machine translation are less useful than evaluations of specific systems by particular users, because evaluation depends very much on the needs of the specific user. A researcher who needs to comb through a vast mountain of information may find rough translations of abstracts useful for tracking overall trends in research or selecting articles for full translation. Nothing less than absolute accuracy will satisfy a lawyer working on a legal brief or a politician whose words will be quoted by the media. Experience shows that the more a user uses a system, the more efficient the work. A major obstacle to machine translation development is the reluctance of developers to provide detailed information about performance characteristics and to exchange information about their experiences. Potential users are well advised to conduct systematic comparisons on texts of their own choosing.

4. While more than 30 years of development work in a variety of language pairs has not produced a machine that replaces the human translator, today machine translation and related technologies are increasingly seen to augment the work of human translators. Most of the Japanese to English machine translation systems now in use in Japan today succeed because they are limited to particular domains. From a Japanese perspective, these systems can be used effectively for carefully targeted purposes (such as translation of technical manuals) — and it will be possible to improve these systems if resources are expended in research and development. In addition, there are a variety of machine aids for human translators (dictionaries and other composition tools) that have practical uses today even if the prospect of a fully automated, high quality, low cost, general purpose system is over the horizon. However, as mentioned above, the Japanese to English machine translation systems under development in Japan do not meet all the potential needs of US users for high quality, low cost translations.

The question is what should be done and who should do it? The participants in the symposium shared a range of views on this subject. Skeptics wonder whether there is real demand for more Japanese technical information and question whether the government should place priority here. A contrasting view, however, is based on the recognition that markets sometimes fail to meet the national interest — the Japanese approach has been to consciously create demand. There are other reasons to consider. As a technical challenge, Japanese to English machine translation represents an engineering stretch that may spin off progress in other areas of information technology. In other words, it is a test bed for solving natural language

processing problems. In addition, a focused effort bringing together researchers, developers, and users could provide an experiment for new approaches to collaborative R&D.

If policymakers in the United States decide to support development of machine translation technologies, there are a variety of approaches that could be taken. First, international cooperation in machine translation could focus on the "precompetitive" nature of some research themes; it could also speed dictionary development, text sharing and database construction. An international consortium has been proposed and machine translation has been identified as an area for cooperation under the US/Japan Agreement for Cooperation in Science and Technology. US-Japan cooperation could be particularly meaningful if it led to expanded cooperation with some of Japan's premier corporate programs.

Second, the United States government could take the initiative in developing a US strategy. The basic rationale for such an approach might be that there are no organizations in the US with the proper resources to do this independently. One participant in the symposium proposed a three-stage program that would involve the development of prototype Japanese to English systems by two competing consortia involving government, university and industry researchers. In stage one a steering committee could identify the bottlenecks in development of high quality machine translation systems, explore evaluation criteria, and develop a plan for an organizational approach combining government, industry and university researchers and strategists.

Stage two could focus on the development of two prototype systems by two competing consortia. During stage three, interim products could be developed by subcontractors who could focus efforts on machine translation for a wide variety of domains. (A purely private sector consortium is another possibility, but a number of participants in the symposium said that the government would need to play a role as catalyst.)

Third, initiatives could be taken at the organization or corporation level. Today only the largest US organizations have developed systematic mechanisms for expanding access to foreign technical literature. In order to build a constituency of users, the organization must invest the resources in working with them to tailor search, selection and output to their needs.

Such efforts are not mutually exclusive. It is, however, not clear that international cooperation alone would build a sound US research and production base for a new generation of Japanese to English machine translation systems. If policymakers decide to pursue new initiatives, it will be important to consider ways to reduce the fragmentation that today characterizes to a great extent relationships among the research, development and user communities in the United States.

DR. MARTHA CALDWELL HARRIS

Director
Office of Japan Affairs
National Research Council

Dr. Harris holds an M.A. and Ph.D. in political science with a specialization in Asian Studies from the University of Wisconsin (Madison). She was a foreign research associate at the University of Tokyo, where she conducted research on energy policymaking. Harris has taught courses on Japanese politics, international technology transfer and energy policy at the University of Washington and George Washington University. As project director and senior analyst, she participated in a number of studies on international technology transfer and competitiveness at the Office of Technology Assessment, a Congressional research agency. She has written and lectured on Japanese science and technology policymaking, Japan's role in alliance politics and in the Third World. In the spring of 1988 Dr. Harris joined the National Research Council, the operating arm of the National Academies of Sciences and Engineering, to direct a newly established Office of Japan Affairs.

Mr. VALENTINE. Thank you, ma'am.

Dr. Vasconcellos?

Dr. VASCONCELLOS. Thank you, Mr. Chairman. I'm submitting my formal statement for the record, and with your permission I would like to summarize its contents at this time.

Mr. VALENTINE. Thank you.

Dr. VASCONCELLOS. I will be speaking more about machine translation in general, addressing specific questions that were in the charter. Referring first to the current state of the art in machine translation, the most important thing to be said is that MT is being used on large volumes of text and a broad range of subjects.

Examples, in addition to the Air Force's Foreign Technology Division, which will be reported later, include Canada's METEO, which translates 42,000 pages of weather reports a year, and my institution, the Pan American Health Organization, which processes about 800 pages a year—8,000, I'm sorry—in a wide range of subjects.

The fact that so much material is being machine translated means that a broad variety of linguistic structures, domains, and discourse types are being handled. There are at least 11 commercial MT systems currently in active use for the Western languages. Most of these were U.S. initiated. There are also a handful of non-commercial systems.

For Japanese-English, and English-Japanese, there may be 10 more systems in use—and these are being discussed in detail by other witnesses.

The problems, costs, and speed will differ depending on whether the system is specialized, as in the case of METEO, or general purpose—broad domain, as Dr. Harris called it. With METEO, only 3 percent of the output needs correction by a human posteditor. These rates are higher with the general systems, and the purpose of the translation will also affect the extent of the intervention.

Our experience at the Pan American Health Organization is that MT more than doubles the output of the professional translator who is adept at using it, and it can be much more than that.

Now, as to progress in the last 25 years since ALPAC, I think we're all familiar with the technological aspects. I'm going to mention some key developments that have had to do with changes in translation needs and in attitudes.

Originally, MT was used to capture information in Russian and make it available in English. Later, after the Cold War, the direction shifted and translations were aimed at disseminating information instead of collecting it.

In the push to sell products overseas, MT was enlisted to translate manuals from English into a dozen different languages. Now, however, we are coming full circle. While the need for product documentation continues, we are also seeing a demand for into-English again. Japanese, of course, is the top priority, and there is still interest in Russian, German, and French, as Mr. Bostad will be telling us.

In addition, we have added Chinese, Korean, Farsi, Arabic. And soon to join the list will be languages such as Czech, Hungarian, and Polish. All these languages will be very important in technology transfer in the coming years.

For this kind of application, in which we want to capture information just in general, we must have large general purpose systems. We cannot use specialized systems. For this there are only five strong general MT source languages currently—Russian and English, of course, and also French, German and Spanish.

For a system already in place, development of a new source language is approximately three times more costly than the addition of a target.

MT is being seen at the very least as a good scanning device and, at best, as the ultimate solution to the information bottleneck. Many feel that technology transfer cannot take place without MT. Manual translation is too slow, too expensive, and in too short supply, especially in the non-Western languages, to meet the need for real-time or near real-time translation of large amounts of material.

As to the next question: Is a new assessment needed?

We've heard already that the situation in Japan is in fact being assessed, and rightly so. Elsewhere, however, I think not. MT can only be judged in terms of a system's capacity to be used in a particular application—either directly or with customization. Certainly if any criteria are to be established, this must be done by the developers and users of MT working together.

As to the role of the Federal Government, the government could indeed take a more active interest in MT, as Dr. Harris has indicated. It could serve as a catalyst in developing and supporting a mechanism for the coordinated formulation of policy.

Such a mechanism, while serving—preserving a context of diversity, which I believe is very important, should make it possible to maximize resources by helping to focus investments in areas of highest priority, avoid unnecessary duplication, share information, and identify resources, among other things.

Specifically, the government could help to stimulate funding for the development of needed MT source components. Given that MT development is highly labor-intensive and that the commercial companies have a hard time surviving, the challenge calls for a public-private partnership.

Regarding machine-aided translation, which I understand here to mean mainly lexical databases because we all use word processing, and other software is only peripheral—our productivity with LDBs is even harder to measure than with full machine translation because the situation varies greatly, depending on the language and subject matter, as well as the knowledge that the translator brings to the task. There have been estimates of 7 to 11 percent improvement in productivity.

In the end, whether or not LDBs are productive is secondary to the fact that they do no more than leverage professional translators and, therefore, only address a small part of the language problem. They do not expand the translator pool, and they do not help the monolingual to bridge the language gap.

To speak briefly of our experience at the Pan American Health Organization, I should first point out that in our installation, development and production are carried on simultaneously.

Today, PAHO stands alone among international organizations in using MT as its primary mode of translation. The two systems—

Spanam, from Spanish into English, and Engspan, from English into Spanish, are general-purpose mainframe systems with large and deeply coded dictionaries. They perform a wide range of—they perform on a wide range of subject matter, vocabulary, language style, and discourse types. Data collected for a period of 8-1/2 months showed that MT was used on 79 percent of all jobs into English, and 60 percent of those into Spanish.

Raw or semi-raw translations are sometimes provided for information purposes only. Terminology continues to be added to both systems in response to daily feedback from the translators. The constant exercise of the two systems on production text has led to a substantial improvement in the quality of the raw output.

Thank you.

[The prepared statement of Dr. Vasconcellos follows:]

**MACHINE TRANSLATION:
PROGRESS TO DATE AND SUGGESTIONS FOR U.S. GOVERNMENT ROLE;
PRACTICAL EXPERIENCE AT THE PAN AMERICAN HEALTH ORGANIZATION**

Testimony for the U.S. House of Representatives
Committee on Science, Space, and Technology
September 11, 1990

Muriel Vasconcellos¹

PROGRESS IN THE LAST 25 YEARS AND STATE OF THE ART

The last 25 years have seen slow but certain progress in machine translation, driven by the needs of a rapidly internationalizing world. The major changes are summarized in the sections that follow. Based on these changes, directions for the U.S. Government are proposed. The particular experience of the Pan American Health Organization is summarized in the final section.

Computers and Computation

The period in question has seen an unprecedented revolution in the development of computers and programming languages. At the time of ALPAC (1964-66), the Committee's decisions were predicated on hardware that had less capacity and processing power than a small desktop PC today. Programs were still written in user-unfriendly assembler language. Word processing as we know it did not exist; instead, text input was seen as an insurmountable bottleneck, and editing of the output had to be done by hand and rekeyed by specialized operators. All this was to change dramatically, however, and as the situation unfolded, MT gained impressive new momentum with each step forward--miniaturization, higher-order programming languages, optical scanning, and the mass availability of word processing.

MT technology was originally limited to mainframe operations and therefore to large government or industrial settings, but now it is running on increasingly smaller and less expensive computers and becoming much more accessible to the general public. This is not to say, however, that it is ready for mass-marketing on PCs. A number of factors have held back the porting of heavy-duty MT systems to microcomputers. Only quite recently have hard disks and database management systems had adequate capacity for storing and manipulating the large dictionaries required, and even so, lookup time is still rather slow. Other problems remain to be solved and are being worked on. One of these is customer support, which is labor-intensive and hence incompatible with mass marketing strategies. Also, vendors have been reluc-

¹Chief, Translation and Terminology, Pan American Health Organization, Washington, D.C. 20037.

tant to release large and deeply coded dictionaries that represent years of work, while smaller dictionaries are not sufficiently robust for general-purpose translation. Solutions are being found, however, and the hurdles are falling away. Once the heavy-duty systems are finally ported to PCs, the technology will be accessible to the general public and it will be affordable. Small translation services and independent translators will be able to buy it, and productivity will be directly measurable.

Parallel processing is another recent development that stands to affect MT. This technology, combined with a neural network approach and the capability of programming in declarative language without regard for the order in which operations are carried out, will undoubtedly make inroads into the remaining unconquered territory, but at a high cost in terms of investment in research and reduced availability to the general public.

MT Systems

For the Western languages, there are at least eleven commercial MT systems currently in active use (listed in the order in which they were first implemented a practical product): Systran, Logos, Weidner (now owned by Bravice), Alpnet, Smart, Globalink, Atamiri, Metal, Tovna, DP/Translator, and XLT. It has not been easy for these companies to survive. Long-term investments are needed in order to build powerful general-purpose systems.

Not sold commercially but in practical use are France's B'Vital and Titus, Canada's Météo, PAHO's Spanam and Engspan, and the USSR's Ampar, Nerpa, and Frap.

The problem of Japanese has given rise to 19 systems in varying stages of development. In the following alphabetical list the asterisk (*) indicates that the system is understood to be in practical use: *Atlas-I and II (Fujitsu), *Duet (Sharp), *Hicats (Hitachi), IBM/Tokyo Research, Lamb (Canon), Lute (Nippon Telegraph & Telephone), *Meltran (Mitsubishi), *Medium/Micro (Bravice), *Nippon-Data General, Parole (Matsushita), *Pensée (OKI), *Pivot (NEC), RMT (Ricoh), Smart, SMTR (Resource Sharing), *SWP-7800 (Sanyo), *Systran (Iona), and Taurus (Toshiba).

In addition, there are research systems involving a wide range of languages. In the United States, the academic sector is represented by Carnegie Mellon University, the University of New Mexico, and, under a combined project, City University of New York, New York University, and Monmouth College, while in the private sector there are projects at IBM, MCC, TRW (ECS), and possibly other companies. In addition, there are numerous research activities in Europe.

Quality

Since MT systems are being used at a large number of sites, the next question is whether the quality of the product has improved sufficiently to make it a useful tool. MT quality has always been difficult to measure.

Those who work with the technology have come to the conclusion that it must be judged dynamically in terms of overall system capacity for a given application rather than statically with reference to specific outputs. The factors that enter into defining the "application" include: the purpose of the translation; the language, genre, domain, degree of specialization, repetitiveness, and grammatical structure of the source text; the hardware available; degree of postediting required and availability of personnel for this purpose; and finally, attitudes of those who use it (Vasconcellos 1990). Quality cannot be readily separated from these factors, or from long-term performance. A system must effectively meet the needs of its user, whatever those needs may be.

Two decades ago--when MT was nearly analyzed to death--it was fashionable to speak of "accuracy rates." But between an "accuracy rate" of 80% in 1965 and one of, say, 95% in 1990 lies a difference of such magnitude that only a total view of the situation can begin to tell the story. In the first place, there is no comparison in the volumes of text being machine-translated. In separate testimony we are hearing today about the large volumes and broad range of subjects covered by Systran at the U.S. Air Force Foreign Technology Division. My own institution, the Pan American Health Organization, used MT for 1.8 million words (7,200 pages) of production in 1988 and 2.1 million words (8,400 pages) in 1989 (see final section of this statement). In a more specialized situation, Météo 2 translates 10.5 million words (42,000 pages) of Canadian weather forecasts per year (Chandioux 1989:453) and reports an all-time total of more than 100 million.

The fact that so much material is being machine-translated means that more linguistic structures, more domains, and more discourse types are being handled. General-purpose systems, with large and deeply coded dictionaries developed over the years, are coping across a broad variety of texts. Perhaps the greatest test of how much the systems have improved is that they are responding to linguistic challenges that were not even attempted in the late 1960s. Advances in MT have progressively revealed the true complexity of human language--the messy parts that defy logic such as anaphora, conjoined clauses, ellipsis, metaphor, fragments, unexpressed inferences. Translation programs now address some of these problems to one degree or another, but the effect as measured in the output may not be so visible. No criterion of "accuracy" applied to specific texts is going to reflect the transcendancy of these solutions.

For these reasons, it is now thought best to measure MT quality in terms of a system's capacity to be used in a particular application, either directly or with customization. System power comes from the sophistication and refinement of its linguistic components (analysis (grammar and parser), transfer, and generation), but more than that it comes from the investment of years of work in building up a large base of grammatical and lexical rules through constant exercise of the system on real text. As far as linguistic or AI models are concerned, so far no single theoretical approach has prevailed. Many have been tried, and a number of them have proved effective. Some of the systems are a blend of several approaches, and this also works, because language itself has multiple components that function simultaneously. Again, it is the application that determines the system's effectiveness.

Attitudes

People in general are more willing to accept the shortcomings of MT than they were in the past, and they are more tolerant of raw and lightly postedited output. It may be that the sheer numbers of people involved in the MT field, coupled with the growth in computational linguistics and AI, have had the effect of educating the public.

There is also renewed concern in the U.S. about accessing the large quantities of technical literature being produced in languages other than English--in as many as 20 languages, some say. For information that is already stored in digitized form, MT is being seen at the very least as a good scanning device, and at best, as the ultimate solution to the bottleneck.

Because there is a greater understanding of how complex language really is, there is also increasing recognition that MT will take a long time to be developed. Impatience for quick results and impetuous wholesale rejection of MT have been replaced by an acceptance that the task may take decades and that progress in MT is going to be gradual.

Another change is that development teams are more open and willing to work together than they were in the 1960s and early 1970s. A number of projects are now being carried out in several countries simultaneously. This cooperative spirit is important for the MT field, which relies heavily on people. There is growing recognition, in fact, that at all levels the technology is people-driven: developers need to collaborate closely both among themselves and with users, and long-term success depends on the commitment of translators, editors, and dictionary updaters who work with MT on a day-to-day basis.

Translation Needs

Language Combinations

Originally MT was used to capture technical and scientific information in Russian and make it available in English. However, this direction of events shifted sharply in the late 1960s when U.S. Government funding dried up and at the same time the post-Cold-War climate began to foster translations aimed at disseminating information instead of collecting it. In particular, the growth of computer technology was accompanied by a push to sell products overseas, and MT was enlisted to translate the corresponding manuals and documentation so that they could be launched not only sooner but also simultaneously in multiple markets.

For a system already in place, source language development may be as much as three times more costly than the addition of new targets. Today the only strong MT source languages besides Russian and English are French, German, and Spanish. In contrast, the target languages (mainly from English) include Dutch, English, French, German, Greek, Italian, Norwegian, Portuguese, Russian, Spanish, and Swedish.

While the trend to use MT for product documentation is still on the rise, we are now also seeing a demand for into-English again on the part of scientists and policy-makers. It is not so much translation that they need, but information. Japanese is the top priority, and there is still interest in the traditional sources, i.e. Russian, German, and French. In addition, Chinese, Korean, Farsi, and Arabic are being monitored. To a lesser extent, interest has been shown in Hungarian, Polish, and Czech, as well as perhaps a dozen other languages. These needs contrast with the fact that there are only four languages other than English that have analysis components that are strong enough for general-purpose MT. A strong argument could be made for the development of additional sources, especially non-Indo-European languages. The extent to which existing systems can be enlisted is a complex question that may differ depending on the language and deserves full study.

Specialization

The debate continues over the advantages of specialized versus general-purpose systems. Some experts predict that limited-domain systems will ultimately prevail, since they can promise more reliable results and therefore require less human intervention and long-term development and support than do the general systems. An example of a successful specialized system is Météo 2: only 3% of its output requires intervention by posteditors (Chandioux 1989:453).

General-purpose systems, on the other hand, even though they may not be as consistent in their performance, are directed toward a much broader usership and are therefore more commercially interesting for vendors and venture capitalists. They are absolutely essential if the purpose of MT is to gather information.

Integration

MT now serves in many cases as a link in the larger text-production chain. Publishers expect the hardware and software to be fully integrated. The incompatibility of foreign character sets has been an obstacle which initially impeded transmission via modem and is sometimes still a problem. Another concern is the preservation of typesetting codes so they will not have to be re-entered manually in the target language—a costly process. The challenge has been met at the extrasentential level, but codes embedded within the sentence interact with the translation program itself and are still the subject of research.

Applications

While product manuals will continue to be developed, it's safe to say that there will also be much more use of MT for information purposes, especially in database consultation. Human translation is too slow, too expensive, and in too short supply, especially in priority languages such as Japanese, Chinese, Korean, Arabic, and Farsi, to meet the need for real-time or near-real-time translation of large amounts of material. It is reasonable

to expect that information consumers will find that the unpolished MT product, though sometimes cumbersome to read and susceptible to error, will still give them the competitive edge that they need. There is also the option of light postediting, which can be done by subject-area specialists as well as translators, thus broadening the resource base.

Machine-assisted Translation

The term machine-assisted translation has always been ambiguous. In its accepted sense it refers to traditional human-initiated translation in which the computer is enlisted as an aid, mainly in the form of word processing, lexical data bases, and spelling checkers or text-critiquing software applied to the finished product. It refers to a process in an environment rather than to specific tools. The tools speed up the traditional translator's task to one degree or another. The greatest contribution to productivity is word processing, now used by most full-time professional translators (Vasconcellos 1989a). Spelling checkers and text-critiquing software, though useful, do not contribute directly to the transfer from one language to another. Thus, the only area that would appear to be of interest from the point of view of the language barrier is lexical data bases (LDBs).

Because the situation varies greatly depending on the language and subject matter as well as the knowledge that the translator brings to the task, it is extremely difficult to assess the contribution of LDBs. There have been estimates of 7% to 11% improvement in productivity. On the other hand, many translators have been reluctant to consult LDBs. The main reasons given have included past experience of not finding what they are looking for (the data bases need to be mammoth in order to compensate for this problem), the complexity of the terminological record, and the fact that in large data bases a single term may have many records that need to be sorted through. Workstation environments have been proposed in which terminology is brought up automatically in a window on-screen. The problem with this is that the gloss may not be appropriate for the context. LDBs are caught in a vicious circle: in order to contain the information that translators need, they have to be immense (Eurodicautom and Termium have millions of terms), but they need to be used in order to grow.

Naturally, the glosses contained in LDBs are useful to consult in the preparation of MT dictionary entries, but automatic crossover is out of the question. Efforts to combine terminology data bases with MT dictionaries have never proved fruitful. MT entries require precise semantic and syntactic coding, especially when a word changes its translation depending on the context or when it collocates with another to form an expression. At most, one might envision a flag in the MT input to indicate that a word has a terminological record elsewhere in the system.

In the end, whether or not LDBs are productive is secondary to the fact that they only address a small part of the language problem. They are of only occasional use to non-translators. Their main function is to leverage the experienced professional. They do not add manpower to the translator

pool, and they do not help the monolingual to bridge the language gap.

The largest and best-known of the LDBs are: Eurodicautom, Termium, and TEAM (Siemens). They depend mainly on cooperative agreements for their growth. The German Government also has an important data base, and in Japan there is the Electronic Dictionary Research Institute. A U.S. effort was started at Carnegie Mellon University in the late 1970s but folded quickly. The most useful way to advance the LDB initiative is through the contribution of specialized glossaries and new languages to existing data bases, in exchange for which their administrators are usually willing to offer full access.

ROLE OF THE U.S. GOVERNMENT

There are two basic motivations that should prompt the U.S. Government to take an interest in translation: to monitor foreign technology, and to support the launching of U.S. products in foreign markets. The stakes are high: we do not want to lose the "Information War," and we do not want to lose markets for U.S. technologies overseas. In the translation of product manuals, the private sector has gone a long way toward speeding up the task. Many companies are successfully using MT--Xerox being the prime example. It is in the area of information that the work needs to be done.

There are 63 U.S. Government agencies engaged in monitoring foreign technology. We have seen the important contribution being made by the Air Force's Foreign Technology Division. But there is need for coordination to avoid duplication, and there should be better ways of focusing on the information that is of greatest use. The National Translations Center, which would be in a position to undertake this coordination, is unfortunately slated for extinction. One can only hope that this action will be reversed. However, with or without the Center, machine translation can play a key role as a screening device. This task is so overwhelming that it is totally beyond the capacity of traditional human translation. But translators should take heart from the fact that as more material is identified, more work will be generated for them as well.

The U.S. Government can serve as a catalyst by supporting a mechanism for the coordinated formulation of MT policy. Such a mechanism, while preserving a context of diversity, should make it possible to maximize resources by helping to: focus investments in areas of highest priority, avoid unnecessary duplication, share information, identify resources, develop technical criteria, and possibly, reward excellence.

As part of this role, the Government should identify the languages of highest priority and help to stimulate funding for development of the corresponding MT source components. For information-gathering, attention should be given to Chinese, Korean, Arabic, and Farsi, in addition to Japanese. In time, the development of MT source components for the languages of Eastern Europe may also be warranted.

PRACTICAL EXPERIENCE AT THE PAN AMERICAN HEALTH ORGANIZATION

The Pan American Health Organization began to develop MT for internal use in 1977. The installation is distinctive for the fact that development and production are carried on simultaneously in the same environment and that translators provide constant feedback for system improvement. Today PAHO stands alone among international organizations in using MT as its primary mode of translation.

Spanamtm, which translates from Spanish into English, was first placed in practical use in January 1980 and has been in continuous operation since that time, producing today as much as 175,000 words a month. Most of this output is postedited by professional translators, especially when it is for publication purposes. "Light" postediting is used with internal texts such as trip reports, many of which would probably not be translated otherwise. Work on Engspantm, from English to Spanish, was partially supported by a grant from the U.S. Agency for International Development (AID), which permitted the development of a robust parser, as well as transfer and target generation modules capable of producing context-sensitive translations (León & Schwartz 1986, Vasconcellos & León 1988). Collocations and idiomatic expressions are elicited using a number of different strategies. This grant was motivated by AID's desire to disseminate information on health and agriculture in Third World countries. Engspan has been operational since 1985. The new features of Engspan have been retrofitted onto Spanam, and work continues with the goal of bringing it up to the level of Engspan. Examples of raw output are given in Annexes 1 and 2.

Both Spanam and Engspan are general-purpose mainframe systems with large dictionaries (approximately 63,000 and 55,000 terms, respectively). The tasks addressed are quite varied in terms of subject matter, vocabulary, language style, and discourse type. In addition to medicine and public health, they are used for many other subject areas: agriculture, computer science, law, management, sanitary engineering, the physical sciences, etc. Engspan's jobs are often for wide dissemination in Spanish-speaking Latin America.

After several years of operation coupled with ongoing enhancement of the two systems, it was decided to take an in-depth look at the practicality of using MT on an institution-wide basis. Records kept over a period of 8.5 months in 1987-88 showed that MT was used on 79.4% of all jobs received for translation from Spanish into English and 60% of the volume from English into Spanish (Vasconcellos 1989b). Feedback collected during the period showed 85.1% satisfaction with MT compared with 78.1% for human translation. In view of the study's results, it was decided to establish MT as the principal mode of work, as this would make the best use of available talent and resources while at the same time providing faster service.

Raw or semi-raw translations are sometimes provided for information purposes only. Recently, with the mushrooming of specialized data bases in fields related to public health and agriculture, Engspan has been gearing up to serve as an information tool as well as a support for translators.

Terminology continues to be added to both systems in response to feedback from ongoing operations. The constant exercise of the two systems on production text has led to substantial improvement in the quality of the raw output. Many of the posteditors have commented that the machine's product is easier to work with and requires fewer corrections than it did when they first started. One of the posteditors estimates that her productivity is doubled with machine translation.

Engspan, although not commercially available, is also installed at AID and is used at the International Center for Tropical Agriculture in Colombia, and the International Rice Research Institute in the Philippines on a cost-sharing basis.

References

- Chandioux, John. 1989. "Météo: 100 Million Words Later." In: Proc American Translators Association Conference 1989 (Washington, D.C., 1989), ed. Deanna L. Hammond. Medford, NJ: Learned Information. pp. 449-453.
- León, Marjorie & Lee A. Schwartz. 1986. "Integrated Development of English-Spanish Machine Translation: From Pilot to Full Operational Capability." Technical Report of Grant DPE-5543-G-SS-3048-00 from the U.S. Agency for International Development. Washington, DC: Pan American Health Organization.
- Vasconcellos, Muriel & Marjorie León. 1988. "SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization." In: Machine Translation Systems, ed. Jonathan Slocum. Cambridge, New York, etc.: Cambridge University Press. pp. 187-235.
- Vasconcellos, Muriel. 1989a. "M(A)T in the U.S.A." In: Proc Second North American Translators Congress, Regional Center of North America for FIT (Washington, D.C., 1989), ed. Jean-François Joly. pp. 159-167.
- Vasconcellos, Muriel. 1989b. "Long-Term Data for an MT Policy." In: Literary and Linguistic Computing 4(3):203-213.
- Vasconcellos, Muriel. 1990. "Perspectives on the Assessment of Machine-Translated Output." In: Miscellany on Translation Criticism, ed. Milan Hrala. Fédération internationale des traducteurs. In press.

Chagas' disease and blood transfusion	OK	Enfermedad de Chagas y transfusión de sangre
1. Introduction	OK	1. Introducción
Chagas' disease, or American trypanosomiasis, is a parasitic disease caused by the hemoflagellate protozoa, Trypanosoma cruzi. The infection occurs only in the Americas, where it is widely distributed in the peri-urban or rural areas of the tropical or subtropical countries, from Mexico in the North to Argentina and Chile in the South. It is transmitted to man and other mammals mainly by hematophagous insects of the subfamily Triatominae, the triatomine bugs.	GS PP 18	La enfermedad de Chagas, o «la tripanosomiasis» americana, es una enfermedad parasitaria causada por el protozoo hemoflagelado, «Trypanosoma cruzi». La infección ocurre solo en las Américas, donde se distribuye ampliamente en las áreas peri-urbanas o rurales de los países tropicales o subtropicales, de México en el Norte a Argentina y Chile en el Sur. Se transmite al hombre y otros mamíferos principalmente por los insectos «hematófagos» de la subfamilia triatomíneos, los insectos triatómos.
It is thought that from 10 to 20 million persons in the rural and urban areas of Latin America are infected with T. cruzi. In addition, it is estimated that 65 million people live in endemic areas where the risk of infection exists.	PP 17	Se piensa que de 10 a 20 millones de personas en las áreas rurales y urbanas de América Latina están infectados por T. cruzi. Además, se estima que 65 millones de personas viven en las áreas endémicas donde existe el riesgo de la infección.
One monograph would not suffice to review comprehensively and in depth all aspects and developments pertaining to Chagas' disease. Therefore, while this report presents an overview of this subject, its major focus of concern will be the relationship between T. cruzi infection and blood transfusion.	OK TU	Una monografía no bastaría revisar integralmente y exhaustivamente todos los aspectos y acontecimientos pertinentes a la enfermedad de Chagas. En consecuencia, mientras este informe presenta un resumen de este tema, su enfoque principal será la relación entre la infección por T. cruzi y transfusión de sangre.
2. Epidemiology	OK	2. Epidemiología

a) Geographical distribution

OK

A recent review on the subject indicates that humans, domestic animals and triatomids infected with T. cruzi are commonly found in Latin America, with the exception of Cuba and the Dominican Republic. On the other hand, except for Belize, Trinidad and Tobago and Guyana, where a few cases of human infection were described, only infected vectors and wild reservoirs or only vectors were found in the remaining countries of the Caribbean and South America.

a) Distribución geográfica

PP 31 TU

Una revisión reciente sobre el tema indica que humanos, animales y triatomas domésticos infectados por T. cruzi se encuentran comúnmente en América Latina, con la excepción de Cuba y la República Dominicana. Por otro lado, salvo Belice, Trinidad y Tobago y Guyana, donde unos pocos casos de infección humana se describieron, solo los vectores infectados y los reservorios salvajes o solo los vectores se encontraron en los países restantes del Caribe y América del sur.

In Mexico, the infection is more common in the Pacific Coast States (from Chiapas to Nayarit), in the Yucatan Peninsula and in certain parts of the Altiplano.

SD OK

In Central America, the endemic is found in Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua and Panama.

PP 19

In Central America, the endemic is found in Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua and Panama. In Costa Rica, the vector is present in the country's central plain and adjoining zones, extending primarily to the northwest and southwest regions. The center of the endemic areas is in the province of Atlajueles. In El Salvador, the endemic occurs in most of the rural areas where the vector is present in 30-80% of the rural homes and small or medium urban nuclei accounting for 70% of the dwellings in the country.

In Guatemala, the endemia is more common in the departments SD SD SD OK. En Guatemala, la endemia es más común en los departamentos de Chiquimula, Jalapa, El Progreso, Santa Rosa and Zecapa. In Honduras, the vector is present in twelve of the departments. Depending on the species, up to 34% of the insects were infected. Early studies from Nicaragua indicated that domestic triatomids are found in the mountainous area of the northwest, the central region and west of the Pacific Coast.

10480 ENGSPAN (TH) IS A COPYRIGHTED WORK OF THE PAN AMERICAN HEALTH ORGANIZATION
 9H080650000000
 CHAGAS DISEASE & TRANSFUSION

	PARSER STATS			
	OK	PP	NO	TOTAL
# OF SENTENCES	15	7	0	22
% OF TOTAL SENTENCES	68	32	0	
AVERAGE LENGTH	10	24	0	21
COMPLETE SENTENCES	11			10
LONGEST SENTENCE	30			
AVERAGE WORDS PARSED		20		
CURRENT ALLOCATION LIMIT		50		
TOTAL WORDS + PUNC				489
TOTAL WORDS NOT FOUND				6
TOTAL AFTER UNIT LOOKUP				460
ACCOUNT TOTAL=				1370620

CONTROL DE LA MALARIA

OK

La grave situación de la malaria en la Región requiere la OK
 acción de todos los niveles de decisión a fin de conseguir
 el desarrollo de programas eficaces de prevención y control
 de la malaria.

La variedad y complejidad de los factores de riesgo no pp 15
 justifican mantener como táctica única de control, el
 rociamiento intradomiciliar de insecticidas residuales
 aplicados en forma indiscriminada.

Por una parte, es indispensable obtener la información pp 37
 pertinente para que los Gobiernos centrales y locales tomen
 la decisión política para el desarrollo y fortalecimiento
 de sistemas locales de salud que incorporen en su plan de
 acción el control integrado de la malaria y que en dicho
 desarrollo, se promueva vigorosamente la asignación de
 recursos humanos y financieros de los distintos sectores
 afectados por el problema.

Por otra parte, son necesarios los ajustes técnicos para NO 07 TU
 que la planificación del control se realice teniendo en
 cuenta el conocimiento epidemiológico que permite
 estratificar el área malarica y analizar los grupos y
 factores de riesgo con el fin de aplicar las intervenciones
 necesarias para la prevención y el control.

CONTROL OF MALARIA

The serious situation of malaria in the Region requires
 the action of all the decision-making levels in order to
 obtain the development of effective programs of prevention
 and control of malaria.

The variety and complexity of the risk factors do not
 justify to maintain as tactical only one of control, the
 household spraying of residual insecticides applied
 indiscriminately.

On the one hand, it is indispensable to obtain the
 pertinent information so that the central and local
 Governments take the political decision for the development
 and strengthening of local health systems that incorporate
 in their plan of action the integrated control of malaria
 and that in said development, is promoted vigorously the
 allocation of human and financial resources of the
 different sectors affected by the problem.

On the other hand, they are necessary the technical
 adjustments so that the planning of the control is carried
 out taking into account the epidemiological knowledge that
 makes it possible to stratify the malarious area and to
 analyze the groups and risk factors for the purpose of
 applying the interventions necessary for the prevention and
 the control.

El reajuste en la planificación de las acciones dentro de 70 PP 40 The readjustment in the planning of the actions within los sistemas locales de salud y el estímulo para lograr la local health systems and the incentive in order to achieve participación social en la prevención y el control de la the social participation in the prevention and the control of malaria will permit the implementation of the strategies of primary health care and the decentralization. for the purpose of that the interventions are carried out in integrated form utilizing all the available resources and in accordance with the epidemiological situation // site.

OK

Para que la reorientación de los programas se pueda PP 31 So that the reorientation of the programs one can carry out is necessary that are produced changes in the policies that define and regulate the actions of the programs and in the strategies both at the overall level and operational that indicate how those actions will be put in practice.

PP 45 The felt need to carry out control measures with a comprehensive approach so that the combat to malaria is successful, requires that the programming activities take into account the new contributions of the biomedical and social sciences and that the epidemiological method constitutes the backbone of this comprehensive approach

enfóque integral

Existe consenso entre los países de la Región en que. NO 09
para obtener mejoría en la situación de malaria, se
requiere llevar a cabo cambios conceptuales y operativos en
los distintos niveles de conducción y ejecución del
programa.

En el proceso de estudio que debe preceder al PP 33 TU

establecimiento de las estrategias para un programa de
vigilancia, prevención y control de la malaria es preciso
determinar la trascendencia social y económica que el
gobierno le concede al problema dentro del plan nacional de

salud y como parte del programa de desarrollo integral del
país. Si la importancia que el gobierno confiere al
problema de la malaria es real, las decisiones
gubernamentales se deberán reflejar en sus políticas

económicas, financieras, educativa y de vivienda rural, y se
desarrollarán criterios específicos para establecer
prioridades en los niveles políticos y técnicos, de acuerdo
con los riesgos de las diferentes áreas geográficas y
grupos poblacionales allí ubicados.

There exists consensus among the countries of the Region
in which, in order to obtain improvement in the situation
of malaria, is required to carry out conceptual and
operational changes at the different levels of conduct and
execution of the program.

In the process of study that should precede the
establishment of the strategies for a program for
surveillance, prevention and control of malaria is
necessary to determine the social importance and economic
that the government grants him to the problem within the
national health plan and as a part of program for
comprehensive development of the country. If the
importance that the government confers to the problem of
malaria is real, the governmental decisions should be
reflected in its policies economic, financial, educational
and of rural housing, and will be developed specific
criteria in order to establish priorities at the political
and technical levels, in accordance with the risks of the
different geographical areas and population groups there
located.

V0380 SPANAM (TM) IS A COPYRIGHTED WORK OF THE PAN AMERICAN HEALTH ORGANIZATION
 007/86/90
 *H0R055999999

MALARIA CONTROL

PARSER STATS

# OF SENTENCES	OK	PP	NO	TOTAL
	3	6	3	12
% OF TOTAL SENTENCES	25	50	25	
AVERAGE LENGTH	13	50	49	40
COMPLETE SENTENCES	2			9
LONGEST SENTENCE	31			
AVERAGE WORDS PARSED		35		
CURRENT ALLOCATION LIMIT		40		
TOTAL WORDS + PUNC				529
TOTAL WORDS NOT FOUND				0
TOTAL AFTER UNIT LOOKUP				482
ACCOUNT TOTAL=	4013			

§(8

32. CONTRERAS CM, GONZALEZ MT, ZARABOZO D, FERNANDEZ
GUARDIOLA A. Cuid Cient CEMESAM, 9:27-40, 1978.

§(8

Los estudios clínicos de sujetos intoxicados con tolueno pp 13
en forma accidental o deliberada, sugieren un efecto
bifásico del disolvente sobre el sistema nervioso central:
un período de excitación seguido de un período tardío y MO 05
prolongado de depresión. El propósito de este estudio fue OK
delimitar estos efectos en las ratas sometidas a
intoxicación bajo diferentes concentraciones de tolueno. pp 24

Los resultados muestran que existen cambios
electrocorticográficos, electromiográficos y conductuales,
y que dependen del tiempo de exposición y de las
concentraciones utilizadas. Se encontró que en los 3 pp 30
aspectos estudiados existen semejanzas entre la
intoxicación con tolueno y la secuencia de alteraciones que
producen los anestésicos volátiles, que también son
sustancias liposolubles. Se discuten los resultados en pp 36
cuanto a dichas semejanzas y en cuanto a la polineuropatía
que se ha descrito en sujetos expuestos a los disolventes
en relación con el período tardío de depresión del sistema
nervioso, que apareció en este estudio.

§(8

32. CONTRERAS CM, GONZALEZ MT, ZARABOZO D, FERNANDEZ
GUARDIOLA A. Cuid Cient CEMESAM, 9:27-40, 1978.

The clinical studies of subjects intoxicated with toluene
accidentally or deliberately, suggest a two-phase effect of
the disolvent on the central nervous system: a period of
stimulation followed by a late period and prolonged of
depression. The purpose of this study was to delimit these
effects on the rats submitted to intoxication under
different concentrations of toluene. The results show that
there exist electrocorticographic, electromyographic and
behavioral changes, and that they depend on the time of
exposure and of the concentrations utilized. It was found
that in the 3 aspects studied there exist similarities
between the intoxication with toluene and the sequence of
alterations that produce the volatile anesthetics, that
also are liposoluble substances. There are discussed the
results in regard to said similarities and in regard to the
polyneuropathy that has been described in subjects exposed
to the disolvents in relation to the late period of
depression of the nervous system, that appeared in this
study.

39. FERNANDEZ-GUARDIOLA A, CONDES-LARA M, JURADO JL.
CALVO JM. Adv Physiol Sci 17:817-824.

8)8

Algunos autores han demostrado que la estimulación del sistema límbico, repetida a baja intensidad (kindling), produce convulsiones. Existen algunas estructuras que cuando se les aplica la estimulación de tipo kindling no necesariamente tienen crisis epilépticas, pero pueden reflejar alteraciones permanentes. El presente trabajo describe: 1) los cambios en el sueño que se producen con el kindling del núcleo dorsal del rapé en gatos con electrodos crónicamente implantados. Los cambios consistieron en una disminución en la latencia del sueño MOR, un aumento de los eventos efásicos y de la atonía durante el sueño de ondas lentas y un aumento en el tiempo total de MOR, y 2) los cambios que produce el kindling del quiasma óptico en los potenciales evocados visuales registrados en el cuerpo geniculado lateral, en la corteza visual y en el electroretinograma // de gatos curarizados. OK
El kindling produjo un aumento de la excitabilidad manifestada, debido al aumento en la amplitud de los potenciales evocados.

8)8

42. FERNANDEZ-GUARDIOLA A, JURADO JL, SOLIS M.

39. FERNANDEZ-GUARDIOLA A, CONDES-LARA M, JURADO JL.
CALVO JM. Adv Physiol Sci 17:817-824.

Some authors have demonstrated that the stimulation of the limbic system, repeated at low intensity (kindling), produces convulsions. There exist some structures that when there is applied them the stimulation of kindling type not necessarily have epileptic crises, but they can reflect permanent alterations. The present work describes: 1) the changes in the sleep that are produced with the kindling of the dorsal nucleus of the raphe in cats with chronically electrodes crónicamente implantados. The changes consisted of a reduction in the latency of the sleep REM, an increase in the events fásicos and of the atonia during the sleep of slow waves and an increase in the total time of REM, and 2) the changes that produces the kindling of the optic chiasm in the visual evoked potentials registered in the lateral geniculate body, in the visual cortex and in the electroretinogram // of curarized cats. The kindling produced an increase in the manifested excitability, owing to the increase in the amplitude of the evoked potentials.

42. FERNANDEZ-GUARDIOLA A, JURADO JL, SOLIS M.

Rev Med Mex 57:76-85, 1978.

SJS

El estudio incluyó a 10 voluntarios pagados, entre los 19 PP 16 y los 21 años, y, a excepción de uno, estudiantes. Cada OK individuo tomó parte en cuatro sesiones experimentales. La OK primera fue considerada como habituación; la segunda, MO 03 control; y en la tercera y cuarta se administró triazolam PP 15 oralmente en cantidades de 0.25 y 0.5 mg, respectivamente, 30 min antes de iniciar la sesión.

En esta prueba, los individuos se mantenían en estado de PP 28 GT vigilia durante un tiempo extenso (1 hora), respondiendo a estímulos visuales y, simultáneamente, se evaluaba la actividad eléctrica cerebral y los signos periféricos, tales como la frecuencia cardíaca, la frecuencia respiratoria y la respuesta galvánica de la piel.

De acuerdo con los datos descritos, el triazolam causó un OK alto grado de sueño. Este también fue la impresión PP 11 subjetiva de los individuos, quienes además reportaron como efectos colaterales un ligero vértigo, sequedad de boca y problemas en la acomodación visual.

Sin embargo, consideramos que el triazolam fue tolerado a MO 08 las dosis en que se administró y mostró tener propiedades hipnóticas.

Rev Med Mex 57:76-85, 1978.

The study included 10 volunteers paid, between the 19 and the 21 years, and, with the exception of one, students. Each individual took part in four experimental sessions. The first was considered habituation; the second, control; and in the third and fourth there was administered +triazolam orally in quantities of 0.25 and 0.5 mg, respectively, 30 min before initiating the session.

In this test, the individuals were maintained in state of wakefulness during an extensive time (1 hour), responding to visual stimuli and, simultaneously, there was evaluated the cerebral electric activity and the peripheral signs, such as the heartbeat, the respiration rate and the galvanic skin response.

In accordance with the data described, +triazolam caused a high degree of sleep. This also was the subjective impression of the individuals, who in addition reported as side effects a light giddiness, dryness of mouth and problems in the visual accommodation.

However, we consider that +triazolam was tolerated at the doses in which there was administered and showed to have hypnotic properties.

V0388 SPANAM (TM) IS A COPYRIGHTED WORK OF THE PAN AMERICAN HEALTH ORGANIZATION
 05/08/99
 MENTAL HEALTH ABSTRACTS
 -H0R55500000

	PARSER STAYS			
	OK	PP	NO	TOTAL
0 OF SENTENCES	8	10	4	22
% OF TOTAL SENTENCES	36	45	18	
AVERAGE LENGTH	14	31	29	24
COMPLETE SENTENCES	8			17
LONGEST SENTENCE	28			
AVERAGE WORDS PARSED		22		
CURRENT ALLOCATION LIMIT		40		
TOTAL WORDS + PUNC				534
TOTAL WORDS NOT FOUND				1
TOTAL AFTER UNIT LOOKUP				533
ACCOUNT TOTAL-				40638

MURIEL VASCONCELLOS: BIOGRAPHICAL SKETCH

Muriel Vasconcellos holds the B.S. in languages (1958) and the M.S. (1982) and Ph.D. (1985) in linguistics from Georgetown University. She has devoted her entire career to translation, including 18 years specifically to machine translation (MT). She began as a staff assistant on the Georgetown University Machine Translation Research Project (1955-60), where she was involved in many of the major milestones, including preparation of the first proposal for funding and, later, Congressional hearings on the subject (May 1960). She then turned to "human" translation, working first at the Organization of American States (1961-69) and then at the Pan American Health Organization (PAHO) (1969-present).

In 1977 Dr. Vasconcellos was placed in charge of the machine translation program at PAHO, which began production MT from Spanish to English in January 1980. With Marjorie Leçn, she was co-principal investigator under a two-year grant from AID (1983-85) to assist in the development of an English-to-Spanish system. The result of this research, ENGSPAN, has been in operational use since 1985. Her duties were expanded in 1989 when she was appointed head of the Translation and Terminology service.

At Georgetown University, Dr. Vasconcellos was an adjunct lecturer on translation (1976-78) and then developed and taught the course "Machine Aids to Translation" (1981-88). She was co-principal investigator under a grant from NIH to do annotated translations of historical Brazilian documents (1977-79). She serves on the editorial board of the journals Language International, Jerome Quarterly, and Machine Translation and has published more than 50 articles on translation and machine translation. She is active in the American Translators Association (ATA), for which she edited the book Technology as Translation Strategy (ATA, 1988) and serves as chair of the Committee on Translation and Computers (1987-present). She is also a member of the The American Association of Language Specialists, the National Capital Area Translators Association, the Washington Linguistic Society, and the Linguistic Society of America. She is currently active in the establishment of the International Association for Machine Translation and its regional section, the Association for Machine Translation in the Americas.

Mr. WALGREN. (presiding) Thank you very much to both of you; we appreciate your testimony.

Let me ask, what are the considerations when you think of a national program as opposed to an international program? Could either of you just address that question briefly?

Dr. HARRIS. Yes, I can, Mr. Chairman, I should say that—I can give a few comments, Mr. Chairman. I should say that our symposium did not go into great detail on this subject. But—and I know that my colleague, Dr. Vasconcellos—can update us on the status of the international consortium that's been initiated recently.

But I would say that there are some areas in which it would appear that strengthening of the American base would require an American effort, and I would think particularly with regard to development and production here in the United States.

There's certainly many areas where international cooperation could play a role in production of dictionaries and sharing of source text, et cetera. But it would seem that there are others where even if such an international project were undertaken, that there might not be a result of an effort here within our own country to actually develop and manufacture and market MT systems. But I would ask my colleague Dr. Vasconcellos what her views are.

Dr. VASCONCELLOS. I see international cooperation as being a—very important to avoid duplication, to identify resources, to have colleagues working more closely together than they have been in the past. Currently, we're working with Professor Nagao in Japan on the establishment of an international association for MT.

We are also starting an American association. It will be actually for the Americas—because it will include Canada—so that we can focus on our particular needs in the Western Hemisphere within the framework of the international association. I don't see this initiative as taking the place of industry or government participation in areas of national interest to the United States. But I do see that there's an important—it will give focus to the issues and possibly save resources, and sharpen our approaches. So I think that that's very important.

Mr. WALGREN. So it's largely a question of avoiding duplication and stimulating each other, but you have to develop something to bring to the table in any international cooperative effort.

Is there a way to describe what we're doing now that we could bring to that table? Are there measures of the present effort in pursuing machine translation? Does it translate into researchers funded by given dollars, or departments focused on this?

How do you measure the present U.S. effort to have something to bring to the table?

Dr. HARRIS. I would think that the best way to answer your question, Mr. Chairman, is to think of the expert witnesses you've assembled for the hearing this morning. They represent some of the premier research institutions in our country. I won't go down the list, but I think it will become clear as the morning—as time goes by—that we have great strength in some of our university programs, particularly in the development of theory.

The question, however, becomes the role of industry in development and application. And here, of course, there are some American companies that are involved. But I would say that our great

strengths right now may be more in the area of research and linguistic theory, and other areas that are certainly by natural language processing absolutely necessary for building the foundation for machine translation systems.

Mr. WALGREN. The—running through the testimony there's a theme of "this is a long-range project", with high costs involved. I gather those costs are not in the theory side of it then, but in the—in the application of the theory. Is that—

Dr. VASCONCELLOS. That was what I was going to say. What we can bring to the table is a fund of experience that we've had with other language combinations. And in the long run, it's only through exercising systems and building dictionaries over time, based on solid theory, that we're going to be able to solve the problem. There are going to be no quick breakthroughs. It's a long, slow process.

I don't know whether it's necessarily a costly process, but it's going to be based on exercise of real text.

Mr. WALGREN. Is this true even with technical information? Because my instinct is the technical information is a little bit more finite in scope than literature as a—in a novel form or an art form.

It would seem the technical information has certain parameters that repeat themselves and that it would allow for a given area to be mastered without having to have a complete automated capability understanding all the nuances of language in its use. Is that true, or not true?

Dr. VASCONCELLOS. I agree—the nuances of language would keep us all busy for the rest of our lives; and certainly technical literature is more finite. Nevertheless, it still has a lot of problems.

I think Dale Bostad, when he testifies about his work with the Air Force, can tell you he's—they've been now for 12 years being quite viable translating from Russian, but it took a long time to get to that point. And if we're going to capture information, we can't zero in on specific technical areas; we have to have immense dictionaries with all the technical terminology to cover all the fields that we're trying to catch up on.

So I'd still see it as a very large project, but certainly not as difficult as other kinds of language, as you say.

Mr. WALGREN. Let me ask—and you've touched on it, Dr. Harris—why the Japanese are putting such a significant effort into translating from Japanese into English? What are the direct benefits that they see?

Dr. HARRIS. I guess I could answer that in a number of ways. First of all, I think that the Japanese see English as an international language, both in science and technology, and in global markets. And so in order to operate efficiently around the world, it's almost imperative for their companies to have the capability to fairly quickly translate from Japanese into English.

And what we're hearing—what we heard at the symposium—is that for many companies operating internationally, translation may actually become a bottleneck for the introduction of new products. Unless you can produce the technical manuals in timely fashion, the translation may actually hold up your marketing effort. So I think that's one reason.

But I would look more broadly. It appears that the Japanese industry, research community, government as well, sees machine translation as a kind of a—almost a symbol for a new type of challenge in information technology. And they're using this as, I think, one of a number of examples of focused R&D efforts which they expect may have spinoffs in the applications in a number of fields.

So I think there are immediate reasons, as well as some broader ones, for their efforts. And, quite frankly, the resources of these large electronics companies allow them to look over the long run to technologies that may play a major role in competitiveness a few years down the road.

Mr. WALGREN. All right. Let me just, then, touch on one other point that you've also raised, and that is the role of the government in the United States as a catalyst.

Is it clear that the government alone would not be the proper entity to accomplish this? We do everything in terms of joint research with industry and universities and government, some direct research and some organizing role, some funding roles.

How heavily should we be looking to see the private markets, or proper entities, driving this in our country before we feel that it's right to go forward with a government effort?

Should we be waiting for the private sector to do this as opposed to the government trying to make it happen?

Dr. VASCONCELLOS. The private sector could never sustain it. The MT companies have been floundering for years. It's very anti-economic for them. And even if you take a large company which has embedded in it an MT project, it runs at a loss for the company, which is the case already in Japan.

That's why we're concerned; we've seen how these companies have tried and just been unable—the nature of the task is so challenging and goes over such a long term that customers—a customer base cannot normally support this.

Mr. WALGREN. Do you have any estimates of what a significant level of participation by the United States would be in an effort to develop a Japanese to English capability?

Dr. HARRIS. I take it you mean a funding level or something of that nature. I'm not sure that I'm the best qualified to speak to that point, but I can mention one estimate that was presented at our symposium. A Japanese expert on this subject mentioned that a machine translation development effort on a mainframe computer would probably involve 50 to 70 people over a five-year period or so; and that would just be the initial effort. He estimated it at 13 to \$15 million worth. So we're talking about quite a substantial amount of funding, and that just for really the preliminary stage.

Dr. VASCONCELLOS. Japanese, of course, is a tremendous problem. It's one of the hardest languages to analyze automatically that we've approached so far. On the other hand, our Spanish—our English-to-Spanish and Spanish-to-English programs over 12 years have not really topped more than 2 million; and we have very effective systems. So there may be a place in between the two extremes.

I would like to also emphasize that the whole idea of coordinating and avoiding duplication is, I think, one of the keys to the problem. And just on another point, the government—the House Appropriations for the Legislative Branch—has recently slated the Na-

tional Translation Center for extinction on a sum of \$127,000. A center like that, for example, could coordinate and go a long way—has plans to pinpoint the kinds of technology that people are needing, and to identify where translations are available, or where translations are needed.

And I think that the whole effort needs to be much more closely coordinated—and with small investments we can do that—before we go into a large project, I would think. We need to see where we're going.

Mr. WALGREN. Where was that center located?

Dr. VASCONCELLOS. What?

Mr. WALGREN. Where was that center located, in what—

Dr. VASCONCELLOS. It's in the Library of Congress, in—

Mr. WALGREN. In the Library of Congress. And they had \$127,000.

Ms. VASCONCELLOS. \$127,000, and it's being scratched right now, unfortunately. And it was performing a very valuable service. And this is the kind of service that we need—we need to coordinate very badly.

Mr. WALGREN. I see. Well, all right.

Well, let me, on behalf of the Committee, thank you for your testimony. We appreciate your being a resource to us in this area and look forward to talking with you about it in the future.

All right, the next panel is Mrs. Deborah Wince-Smith, who is the Assistant Secretary for Technology Policy with the Department of Commerce, and she is accompanied by Dr. Joseph Clark, who co-chairs the Task Force on Access to Scientific and Technical Information, and is a Deputy Director of the National Technical Information Service.

Also on this panel will be Dr. Charles Brownstein, who is the Acting Assistant Director of the Directorate for Computer and Information Science and Engineering with The National Science Foundation, accompanied by Dr. Y.T. Chien, who is the Division Director for Information, Robotics, and Intelligent Systems for NSF; and also Dale Bostad, Technical Advisor of—for the Directorate of Translation with the Foreign Technology Division of the United States Air Force.

Well, welcome all of you to our process. We appreciate especially your coming, Mrs. Smith. We look forward to the future in many ways and wish you the very best with that. But let's have your testimony first and then we'll go through the panel in the order in which I introduced you to the record.

STATEMENTS OF DEBORAH L. WINCE-SMITH, ASSISTANT SECRETARY FOR TECHNOLOGY POLICY, DEPARTMENT OF COMMERCE, WASHINGTON, D.C., ACCOMPANIED BY JOSEPH E. CLARK, CO-CHAIR OF TASK FORCE ON ACCESS TO SCIENTIFIC AND TECHNICAL INFORMATION, AND DEPUTY DIRECTOR, NATIONAL TECHNICAL INFORMATION SERVICE, DEPARTMENT OF COMMERCE; CHARLES BROWNSTEIN, ACTING ASSISTANT DIRECTOR, COMPUTER AND INFORMATION SCIENCE AND ENGINEERING, NATIONAL SCIENCE FOUNDATION, WASHINGTON, D.C., ACCOMPANIED BY Y.T. CHIEN, DIVISION DIRECTOR, INFORMATION, ROBOTICS, AND INTELLIGENT SYSTEMS; AND DALE BOSTAD, TECHNICAL ADVISOR, DIRECTORATE OF TRANSLATION, FOREIGN TECHNOLOGY DIVISION, U.S. AIR FORCE, WRIGHT PATTERSON AIR FORCE BASE, OHIO

Mrs. WINCE-SMITH. Thank you, Mr. Chairman. I very much welcome the opportunity to testify today on the status of Japanese to English machine translation. You very kindly have introduced Dr. Joseph Clark, and I'd also like to introduce Dr. Phyllis Genther, who has just recently joined the staff in the Department of Commerce as the head of our Japanese Technology Programs, including the Japanese Technical Literature Program.

Mr. WALGREN. Welcome.

Mrs. WINCE-SMITH. By convening this hearing, it is clear that you and the other members of this Subcommittee are keenly aware of the important role we expect machine translation to play in our efforts to increase U.S. access to foreign scientific and technical literature.

Our particular expertise in the Commerce Department lies in the field of Japanese to English machine translation so I'd like to focus most of my remarks on that area.

Actions taken to implement the 1988 head of government U.S.-Japan Agreement on Science and Technology clearly underscores the importance the Administration places on scientific and technical information.

This Agreement, as we know, calls for improving access to, and for the widest possible dissemination of, Japanese scientific and technical information.

The Task Force formed under the Agreement has agreed that one possible way to accomplish this goal is through international cooperative projects on machine translation.

In order to compete in the global marketplace that is going to be increasingly competitive in the years ahead, we know our business and academic leaders must have fast and complete access to Japanese government and private sector technological advances. In order to do so, they face an overwhelming amount of literature that they must review, assimilate, and most of all, utilize.

By current estimates, Japan publishes over 10,000 scientific and technical journals each year. This number is in addition to numerous books, technical papers, reports and various forms of grey literature, which do not get published—and in many cases, may be the most strategically important information of all.

And most importantly, all the indicators that the amount of pertinent and available Japanese information will increase very rapid-

ly in the next few years as Japan completes its transformation to an "information based society."

Japan sees the various types of machine translation not just as a method to increase the number of quality translations for their researchers. They perceive it as an important step in learning more about natural language processing technology in general, which could be one of the key enabling technologies of the next century with widespread effects that cross industry sectors with multiple applications. And I think that is clearly one of the reasons why we see such a large scale private sector involvement in these efforts in Japan, that they don't perceive machine-aided translation as limited to translation alone.

Equal to the problem posed by the sheer volume of literature is the language barrier that we are all so familiar with. Nearly all our researchers and executives run squarely into the language problem which prevents them from skimming Japanese documents or having someone readily available to whom to turn. The relative scarcity of Americans with Japanese language capability also keeps translation costs as high as \$100 per page of a double-spaced document. And I understand that right now in the United States there are only 200 certified Japanese to English translators.

And finally, this creates a backlog of information that cannot be translated in a timely manner. We need real time access to Japanese breakthroughs.

The Modern Language Association estimated that while 23,454 American students began Japanese language studies in 1986, the attrition rate was nearly 50 percent per semester, leaving only about 300 by the fourth year. Therefore, even when the present group of language students is trained, we expect that most Americans will have to continue to read translations of Japanese technical literature rather than rely on the original documents.

We in Commerce believe machine translation offers one possible longer term solution to the dilemma for U.S. researchers and business executives posed by these problems.

Under the auspices of the Japanese Technical Literature Act and through our leadership role on the U.S.-Japan Scientific and Technical Information Task Force, we are studying Japanese to English machine translation as one part of our effort to disseminate knowledge and increase access to Japanese technology.

At the same time, we have a complementary effort under way that will help ensure increased interest in using the materials which could be generated by a good Japanese to English machine translation system. It will not matter how much information we make available if U.S. researchers and, again, the corporate executive hierarchy in our businesses, do not recognize its utility.

We are making great efforts to educate Americans on incorporating Japanese scientific and technical information into the everyday life of their research organizations and companies. This year we will be holding conferences targeted toward corporate executives and information managers as opposed to the traditional R&D community.

Some companies are well ahead also in pursuing this strategy, and I'd like to mention that MCC—the consortium in Texas—has a very active program in this area, as does Motorola and the private

research institute of Battelle in Columbus. I was recently out in Battelle and learned that they target corporate executives with a one-page summary of key foreign developments, realizing that someone at that level will never pay attention to anything that's really beyond a couple of paragraphs in length.

In the Commerce Department's July 1988 Report to Congress on the "Current Status of Japanese to English Machine Translation," we stated that little interest existed within the United States Government in machine translation since the 1965 ALPAC report concluded that the cost of funding a viable system would be prohibitive and that human translation was a better alternative. The background of what happened is recounted in an upcoming article to be issued in the most current volume of the National Science Foundation's "Insight" publication—and we in Commerce provided the key background information to that important article.

Why is the interest in machine translation stirring after a hiatus of 25 years?

We in Commerce believe there are several factors.

1. The cost of computing has dropped enormously, while the cost of human translation has risen enormously, especially for non-Roman alphabet languages. It is likely that machine-aided translation rather than machine translation per se will be the most useful and cost-effective in the short term, not necessarily in the long term.

2. Evolutionary work has taken place on computational linguistics and language theory.

3. There is a growing infrastructure of translation software and machine and machine-aided translation dictionaries, which we've heard about this morning. These dictionaries and programs are more and more capable of resolving huge numbers of ambiguities encountered during translation. For example: Is a chair an object to sit on or a person who runs a meeting?

4. The information explosion requires a method by which U.S. researchers can determine quickly what is and what isn't of value in foreign technical literature.

5. Machine translation is an important step in learning about subjects such as natural language processing, which may become important future generic enabling technologies that are not industry-specific but will have broad applications throughout our industrial product and service base. And I particularly want to emphasize that point because I think it has relevant impact on how the government may respond to the machine translation challenge.

6. Finally, Japan has become a major economic competitor in the world marketplace, and in parallel has become a very key source of new and innovative research and development.

I will leave the technical explanations about U.S. machine translation efforts to our other experts—Dr. Vasconcellos has testified already and we're looking forward to hearing Dr. Jaime Carbonell as well as my colleague from the National Science Foundation.

Nonetheless, let me say, some form of machine or machine-assisted translation is possible and inevitable.

What have we done, and where are we going in the Administration?

In order to address that question, it is necessary to view machine translation within the context of our overall program to increase access to, and U.S. awareness of the need to track and, most importantly, utilize Japanese science and technology.

Most of the agencies with representatives on the various task forces under the 1988 Agreement are making very valiant efforts to increase our access to the Japanese science and technology base. And each of the units in our new Technology Administration in the Commerce Department contributes to this general effort; all of which are underscoring a fundamental premise of the Agreement that was signed in the previous Administration, namely, that in the United States we face a major imbalance in the flow of knowledge and technology between the United States and Japan, which we must devote all efforts to redress.

In Commerce, we conduct our activities mandated under the Japanese Technical Literature Act through our new Japan Technology Program. Congress, including some of the current Members of this committee, recognized when they passed the Act in 1986 that the United States could no longer afford the luxury of the not-invented-here syndrome, and we must be receptive to technological developments wherever they occur.

Under the Japanese Technical Literature program our activities are including a quarterly newsletter, annual directories, sponsorship of the JTEC missions, one-on-one counseling sessions, studies and special conferences, such as I've alluded to earlier.

The activities of our Japanese Technical—Technology—Program are closely coordinated with, and complemented by, the efforts of the National Technical Information Service to collect over 1,500 Japanese technical documents annually.

Finally, the National Institute of Standards and Technology provides valuable assistance in preparing special reports and advances on Japanese technologies such as advanced materials, intelligent buildings and superconductivity published by the Japan Technology Program.

And I'm sure many on the Committee, as well as in our government and outside, are aware of the report that Secretary Mosbacher issued this spring on Emerging Technologies that provided an assessment of where we stand vis-a-vis our competitors in Japan and Europe, and four broad areas, including information, technology materials, the life sciences, and manufacturing.

NIST experts travel to Japan and around the world to assess important developments and report their findings back. This Thursday we're going to be having a meeting with U.S. industry researchers and executives on Japanese work and state-of-the art in diamond thin films; and this is an example of the type of mission that we're pursuing out of the National Institute of Standards and Technology, and the way to disseminate to the key users technology developments.

As we heard this morning from Dr. Harris, the Department of Commerce sponsored a major symposium at the National Academy of Sciences National Research Council on Japanese to English Machine Translation. This symposium was very significant in that it provided a forum where participants from government, academia,

and business could discuss the current state of Japanese to English machine translation, market prospects, and user needs.

The Academy's has published an excellent report highlighting the major points and conclusions of the symposium.

I would like to emphasize that this symposium clearly has helped to revitalize interest in machine translation in the United States, and I believe has set the stage for the Administration to carefully review its recommendations and begin the process of exploring options for the future.

Our involvement in Commerce since the symposium has been reinforced by the April 1990 U.S.-Japan Task Force Report on Scientific and Technical Information, which Dr. Clark chairs. That report, in a bilateral sense, recognizes the key importance of machine translation. It is now providing an ongoing interagency, as well as the bilateral forum, for discussions of possible specific international collaboration in machine translation between the United States and Japan.

The next step is to assess the state of the art in machine translation in Japan. And we will do this by sponsoring, along with the National Science Foundation and the Department of Defense, a Japan Technical Evaluation Center Study Team scheduled for late November to go to Japan. And I understand a number of the members of your panels today will be participating in that activity, including Dr. Jaime Carbonell of Carnegie Mellon, who will be leading the team. It will be composed of top U.S. experts on machine translation—again, drawn from academia, the private sector, and government.

The recommendations of this Study Team, along with those of the Academy symposium and any that might result from this hearing, will again set the framework for us to begin to develop what I hope will become a national strategy on how to use machine and machine-assisted translation to increase the flow of Japanese information to U.S. researchers and our business community.

Any delineation of the final steps we need to reach our goal of a viable machine translation system awaits a final assessment of recommendations from these groups.

Let me conclude, Mr. Chairman, by reiterating that the Administration and the Department of Commerce do believe machine and machine-assisted translation is one important tool that potentially can enhance the competitiveness of American business by ensuring adequate access to information in what is surely becoming an information-based world. And as I mentioned earlier, may have broad applications to a whole new range of products and services in the information arena that we cannot envision yet today.

Thank you very much.

[The prepared statement of Mrs. Wince-Smith follows:]

STATEMENT OF
ASSISTANT SECRETARY DEBORAH L. WINCE-SMITH
BEFORE THE HOUSE SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY
SEPTEMBER 11, 1990
STATUS OF JAPANESE-TO-ENGLISH MACHINE TRANSLATION

Mr. Chairman, I welcome the opportunity to testify today on the status of machine translation. By convening this hearing, it is clear that you and the other members of this subcommittee are keenly aware of the important role machine translation can play in our efforts to increase U.S. access to foreign scientific and technical literature. Our particular expertise lies in the field of Japanese-to-English machine translation and in accordance with the chairman's letter, I will focus on that aspect.

Actions taken to implement the 1988 U.S.-Japan Agreement on Cooperation in Research and Development in Science and Technology underscore the importance the Administration places on awareness of foreign scientific and technical information. The Agreement calls for improving access to, and for the widest possible dissemination of, scientific and technical information. The U.S.-Japan Scientific and Technical Information Task Force formed under the Agreement has decided that one possible way to accomplish this goal is through cooperative projects on machine translation.

Uses For Japanese-to-English Machine Translation

In order to compete in a global market, our business and academic leaders must know about foreign companies' technological

advances. However, they face an overwhelming amount of foreign literature that they must review. By current estimates, Japan publishes over 10,000 scientific and technical journals each year. This number is in addition to numerous books, technical papers, reports and various forms of grey literature. And, most importantly, all the indicators show that the amount of pertinent and available Japanese information will increase exponentially in the next few years.

I should add that Japan sees the various types of machine translation not just as a method to increase the number of quality translations for their researchers. They perceive it as an important step in learning more about natural language processing technology in general, which could be one of the key technologies of the next century with widespread effects.

Equal to the problem posed by the sheer volume of literature is the language barrier. Nearly all our researchers and executives run squarely into the language problem which prevents them from skimming Japanese documents or having someone readily available to whom to turn. The relative scarcity of Americans with Japanese-language capability also keeps translation costs as high as \$100 per page of a double-spaced document. Finally, it creates a backlog of information that cannot be translated in a timely manner. We need real time access to Japanese breakthroughs.

The Modern Languages Association estimated that while 23,454 American students began Japanese language studies in 1986, the attrition rate was nearly 50 percent per semester leaving only

about 300 by the fourth year; the year in which an adequate reading ability finally is achieved. Even more training is needed to translate advanced technical materials. Therefore, even when the present group of language students is trained, we expect that most Americans will have to continue to read translations of Japanese technical literature rather than the original documents.

We believe that machine translation offers one possible longer-term solution to the dilemma for U.S. researchers and executives posed by these problems. The Commerce Department, therefore, under the auspices of the Japanese Technical literature Act and through our work on the U.S.-Japan Scientific and Technical Information Task Force, is studying Japanese-to-English machine translation as one part of our effort to disseminate knowledge about, and increase access to, Japanese technology. It potentially can increase our ability to greatly enhance the timely flow of scientific and technical information, and to track Japanese technological capabilities.

At the same time, we have a complementary effort underway that will help ensure increased interest in using the materials which could be generated by a good Japanese-to-English machine translation system. It will not matter how much information we make available if U.S. researchers and executives do not recognize its utility. Therefore, we are enhancing our efforts to educate Americans on accessing and factoring Japanese scientific and technical information into their research organizations and companies by establishing systems to collect and disseminate

information. For example, we will hold conferences directed toward corporate executives and information managers. We will continue to support JTEC study missions and to increase the number of our newsletter subscribers.

Why The Time Is Right

In the Commerce Department's July 1988 Report to Congress, "Current Status of Japanese to English Machine Translation," we stated that little interest existed within the United States Government in machine translation since the 1965 ALPAC (Automatic Language Processing Advisory Committee) report concluded that the cost of funding a viable system would be prohibitive and that human translation was a better alternative. The background of what happened is recounted in an article on Japan's state-of-the-art in machine translation technology in the most recent issue of Insight published by the National Science Foundation.

I am glad to be here today to report why more interest is starting to stir within the U.S. Government and the private sector. Why is interest stirring after a hiatus of 25 years? There are several factors:

- o The cost of computing has dropped enormously, while the cost of human translation has risen enormously, especially for non-Roman alphabet languages. It is likely that machine-aided translation rather than machine translation will be the most useful at least in the short-term.

- o Evolutionary work has taken place on computational linguistics and language theory.
- o There is a growing infrastructure of translation software and machine and machine-aided translation dictionaries. These dictionaries and programs are more and more capable of resolving huge numbers of ambiguities encountered during translation. Is a chair an object to sit on or a person who runs a meeting?
- o The information explosion requires a method by which U.S. researchers can determine quickly what is and what isn't of value in foreign technical literature.
- o Machine translation is an important step in learning about subjects such as natural language processing, which may become important future technologies.
- o Japan has become a major economic competitor in the world marketplace, and is increasingly a source of new and innovative research and development.

I will leave the explanation of how U.S. machine translation efforts technically compare with those of Japan and other nations, and the details of recent technological advances to technical experts such as Dr. Muriel Vascoucellos and Dr. Jaime Carbonell who

are testifying today. Suffice it to say that we are starting to see light at the end of the tunnel; some form of machine or machine-assisted translation is possible and inevitable.

What We Have Done and Where We Are Going

In order to understand the Commerce Department's interest in machine translation, it is necessary to view machine translation within the context of our overall program to increase access to, and U.S. awareness of the need to track, Japanese science and technology. Most of the agencies with representatives on the various task forces associated with the implementation of the 1988 U.S.-Japan Agreement are undertaking efforts to increase access to Japanese science and technology. Each of our units within the Technology Administration contributes to this general effort.

First, we conduct activities mandated under the Japanese Technical Literature Act through our Japan Technology Program. The United States can no longer afford the luxury of the not-invented-here syndrome. We must be receptive to technological developments wherever they occur. Accordingly, our activities include a quarterly newsletter, annual directories, sponsorship of JTEC missions, one-on-one counseling sessions, studies and special conferences.

The activities of our Japan Technology Program are closely coordinated with, and complemented by, efforts of the National Technical Information Service (NTIS) to collect over 1,500 Japanese technical documents annually. Dr. Joseph Clark, Deputy Director of NTIS, serves as U.S. chairman of the U.S.-Japan Scientific and

Technical Information Task Force. This task force has four goals: 1) to improve the awareness and understanding of organizations and systems established to improve the use of scientific and technical information; 2) to increase the quantity and quality of scientific and technical information; 3) to reduce impediments, if any, to the flow of scientific and technical information such as through the creation of a Japanese Reprographic Rights Organization; and 4) to increase the translation of scientific and technical information.

Finally, the National Institute of Standards and Technology (NIST) provides valuable assistance in preparing special reports on advances in Japanese technologies such as advanced materials, intelligent buildings and superconductivity published by the Japan Technology Program. NIST experts travel to Japan and around the world to assess important developments and report back their findings.

In December 1989, the Department of Commerce sponsored a major symposium at the National Academy of Sciences's National Research Council on Japanese-to-English machine translation. This symposium provided a forum where participants from government, academia and business could discuss the current state of Japanese-to-English machine translation, market prospects and user needs. The National Academy of Sciences published an excellent report highlighting the major points in symposium. Dr. Martha Harris of the National Research Council already testified today on the details of that report. I would like to state that the symposium has helped to

revitalize interest in machine translation in the United States and that we are carefully reviewing its recommendations.

Our involvement since the symposium in Japanese-to-English machine translation has been reinforced by the April 1990 U.S.-Japan Task Force Report on Scientific and Technical Information that recognizes the basic importance of this technology. This task force is providing an ongoing interagency and bilateral forum for discussions of possible specific cooperative projects in machine translation. The Commerce Department is organizing and chairing the U.S. interagency discussions.

The next step is to assess the state of the art in machine translation in Japan. We will do this by sponsoring along with the National Science Foundation and the Department of Defense a Japan Technical Evaluation Center (JTEC) Study Team tentatively scheduled for late November. JTEC studies provide some of the most credible and definitive assessments of current and emerging trends in Japanese technology. This team, which will be led by Dr. Jaime Carbonell of Carnegie-Mellon University, will be composed of top U.S. experts on machine translation from academia, the private sector and government. The final team has been assembled and will meet at the end of this month to prepare for the trip.

The recommendations of the JTEC Study Team, along with those of the National Academy of Sciences' symposium and any that might result from this hearing, will help us develop a strategy on how to use machine and machine-assisted translation to increase the flow of Japanese information to U.S. researchers and executives. Any

delineation of the final steps we need to reach our goal of a viable machine translation system awaits a final assessment of recommendations from these groups.

Let me conclude by reiterating that the Administration and the Commerce Department do believe machine and machine assisted translation is one important tool that potentially can enhance the competitiveness of American business by ensuring adequate access to information in what is surely becoming an information-based world.

Mr. WALGREN. Thank you, Mrs. Wince-Smith.

Let's go on down the panel. Dr. Clark, are there points you'd like to add at this point?

Dr. CLARK. I think we've—thank you, Mr. Chairman—I think we've put together a very comprehensive and brief statement. We'll be pleased to answer questions.

Mr. WALGREN. Very good.

Dr. Brownstein?

Dr. BROWNSTEIN. Thank you, Mr. Chairman—or Mr. Walgren—for this opportunity to talk with you about machine translation. I have written testimony which I've submitted for the record that deals with the questions that were raised expressly, and I'd like to just make a few comments.

I'm presently Acting Assistant Director of NSF for Computer and Information Science and Engineering. To my right is Dr. Chien, who heads our Division of Information, Robotics, and Intelligent Systems, which funds most of the Foundation's current programs in artificial intelligence and funds research and machine translation and associated technologies.

I previously directed that Division and I have in general for a long time been a supporter of research link translation—I believe it provides worthy research goals—and certainly support NSF participation in funding research in this technology.

That happens also to be the Division which follows on predecessor divisions where most of the effort in both computer translation and manual translation in scientific and technical information exchange took place within the Foundation over some period.

I participated in the recent discussions in the context of U.S.-Japan S&T agreement on machine translation and am in the midst of organizing similar talks with the Esprit Program of the European Commission. And as it happens, by coincidence I visited Professor Nagao in Kyoto last year and also the new laboratories at the Kansai science city where there is a large talking telephone project which is essentially a machine translation project.

So I've—I have some impressions about the area—I don't have the technical expertise that Dr. Chien does, and perhaps if there are technical questions we can refer them there.

I just want to make three points to give you a sense of my impressions. First, getting machines to give fully automated high quality translations—that is, translations of the kind that you would expect from a professional human translator—is still a very hard thing to do. And you have to be aware from what you've heard that it would take a great deal of effort, resources, and some time to create systems that would be capable of doing a good enough job to satisfy most of us.

The need to understand that it's a problem that taxes both hardware and software, that has to do a lot of things in addition to just what you would think of as translating the language—it would somehow take the language in, recognize the patterns that provide the language—whether it's a mark on paper, or something on the airwaves, or a digital symbol—somehow code for the machine to represent the code that we call a language, then manipulate it and map it to another language set, and then reverse the whole process to achieve what you would think of as translation.

The fact that you can purchase inexpensive word and phrase translating computers that will fit in your pocket, and that there are commercial systems that will aid human translators at very substantial rates, is really a testimony to a long line of research and engineering and development that's gone on since the infamous ALPAC report.

Most of the work that's needed along the lines of these technologies I would think of as development appropriate for the private sector rather than research—basic research—appropriate for the Federal Government to support.

Making real progress, though, getting to the point of taking input from a scientist in one language and perfect output in another, as I say, will take research, patience—and it's clear that whoever has these things will gain some real advantages. So I see that there is a substantial Federal role.

That brings me to my second point. Progress in working this problem will take improvements in all components of the computer systems that do the work of processing information. And you should understand that machine translation is a whole systems problem. If you can etch finer lines on better materials, you can design more complex circuits and devices to fit into smaller spaces, you would stand a good chance of making machines that will process more information in less time.

The better you understand a problem, the more cleverly you can manipulate information and the better you can create machines that match the problems, the higher the probability that you can gain some more. And if you can make that machine easy for a person to use, there's a better chance that someone will use it.

So I suppose that—the second point is just you need to recognize that there is no silver bullet solution to the business of machine translation. Advances will result only from working on a relatively broad front of scientific and engineering problems in computing, including specific efforts at machine translation.

I mentioned that I visited ATR, and it was quite an interesting thing—to see this new laboratory at Kansai Science City. It's a facility about the size of IBM-Hawthorne. When I was there it had just opened; it was vast empty laboratories. And one of the first things they actually had set up, or a couple of the things they had set up, were systems by which you could design, and prototype computer chips. The other was this machine translation system.

I was—I had a chance to sort of kick the tires, try it out, say some things into it, have them translated into Japanese.

The fascinating thing was that the hardware—every single piece of it, and it was all sitting there in raw form—something that would recognize you speaking, to something that would put out the spoken voice. It was all built here in the United States. The computer that did most of the crunching was an Alliant; the machine that spoke back to you was from Digital, using something called Dectalk—and there were Motorola signal processing chips, I think, at the front end.

It's a testbed that they are using to develop the software for the component texts, recognizing and translated—translating—a very remitted vocabulary. In that case, the demo I saw was associated with making travel arrangements. And the director of the laborato-

ry pointed out that even such a limited system is something that could be a commercially viable product. If they got the software working, they could certainly, in Japan, build the hardware that would do what this collection of American advanced computers was doing. And they were sure that if they could get something working with a couple of hundred words surrounding the problem of making travel reservations, that it would be something of viable utility for people to use.

So there's somewhat of a different point of view there in the sense of doing very complex machine translation for a very limited marketplace, with an eye towards product. They recognize, as well as anyone in this room I think will tell you that going very far beyond that the general systems is a much longer term problem.

My third point concerns something else entirely, and that is about scientific and technical information, and its dissemination. Machine translation technology would be a small part of the larger issue of information transmission and technology transfer. Among other things going on currently at NSF is a scientific and technical information exchange. We host the U.S. end of the Japanese funded service by which U.S. researchers can gain access to Japanese documents—a document database—of technical abstracts, using international data networks.

All of the abstracts entered into this system are prepared in English by the contributing Japanese researchers as a condition of the grants they get from the supporting organizations.

It's a new system. It's just getting into business, and it's business is growing pretty steadily.

The operations of this system are very interesting to observe. The researchers access the system from all over the country using the NSFNET, by telephone, by mail, and occasionally by walking through the doors. They are assisted by an operator who has the specific expertise to do the searches.

He gets these abstracts, and even though he's not doing—and he speaks Japanese, by the way—a lot of translation, because most of them are in English. All of the old ones are in Japanese, the new ones are in English. In every case he has to fix some of the problems that come across in the text, and they're from all over the boardwalk. There are, for example—and we only know this by guesswork—scanning areas—scanning errors—from optical character readers; and you can pick them out in a fairly classic way where—where “d's” are converted into “cl's”; so that you want to spellcheck and find all these strange words and you can see immediately what you need to replace and make it readable.

Occasionally there are manual transcription errors. You can pick out whole sentences that must have been produced by typing into the keyboard with your hands moved one over to the right, because the code is really pretty easy to detect.

But the most telling ones are where there are whole paragraphs or occasionally whole abstracts of perfectly translated and perfectly useless material, resulting from illiteracy or sloppiness in the native tongue of the original writer. And there's a little lesson in that.

I—When we met with Professor Nagao in Kyoto, he told us that it would be a much easier and reasonable task to perfect machine

translation of Japanese to English technical documents if Japanese engineers were better educated in Japanese in the first place.

And having read the abstracts that we put out from the National Science Foundation, I can assure you the problem is not limited to Japanese engineers—and it's not a trivial one. The adage "garbage in-garbage out" takes on a whole new meaning when you have to recycle every word on the way.

But even worse for information exchange is the fact that most applied researchers, especially engineers at their benches in industry are really not avid readers of technical information in their own languages, let alone in foreign languages.

NSF's operations of a very large technical information service program really withered in the '70s as we learned that the most useful information for that community is extraordinarily time-dependent and is exchanged fairly efficiently by person-to-person in direct contact.

Well, we're not giving up then on scientific and technical information exchange via documents. I still believe, as I said before, that it is extremely important. We're also interested in linking researchers more directly, and supporting the direct exchange of basic research or technical information, particularly the large amount of technical information that is now produced on machines, on computers, and exchanged over networks worldwide.

We've taken steps to ensure that compatible U.S. networking standards are supported among the scientific and technical community in Japan, and we have the research community there to participate in international networking activities that we now lead through NSFNET and will gain further advantage with—to the National Research and Education Network. But that, as you're well aware, is a whole other story.

I just wanted to make the point that all of these kinds of activities are critical to improving the activity of U.S. researchers to the rest of the world.

And I think I'll stop there, and I'd be pleased to answer your questions.

[The prepared statement of Dr. Brownstein follows:]

Testimony of
Dr. Charles Brownstein
Acting Assistant Director
Computer and Information Science and Engineering
National Science Foundation

before the
Subcommittee on Science, Research and Technology
Committee on Science, Space and Technology
United States House of Representatives

September 11, 1990

INTRODUCTION

Thank you for this opportunity to appear before the Subcommittee on Science, Research and Technology and present testimony on machine translation. I am the Acting Assistant Director of the National Science Foundation (NSF) for Computer and Information Science and Engineering (CISE). I've been the Executive Officer of CISE since 1986, and before that I directed NSF's Division of Information Science and Technology. CISE supports most of the research that NSF funds which is relevant to machine translation.

I will discuss what NSF has done to support research in machine translation and related technologies; how this research is related to computing and information technology as a whole; and what new challenges and opportunities lie ahead. I'll give my views on the role of the Federal government in furthering research in this technology. First, however, it is useful to give some context to my remarks.

THE CHANGING CONTEXT OF MACHINE TRANSLATION

Machine translation technology has taken on new significance due to increased global communication and trade, the internationalization of knowledge and information production, and ready access to powerful computers and high speed national and trans-continental computer networks. Social, economic and technological change have, in turn, changed thinking about machine translation technology.

Thirty years ago, machine translation was viewed as a testbed and a critical test for the new field of Artificial Intelligence. It was expected that computers could, in a short time, be programmed to do the work of human translators. This turned out to be false, and attention was diverted to other more tractable problems of automation, and to different

issues in artificial intelligence. Much has been learned and accomplished since then in information processing, and today's workers in the field have more mature views about the technology and its application. Machine translation remains a hard problem, and available systems offer partial but useful solutions.

A recent report of the National Research Council, "Japanese to English Machine Translation", (National Academy Press, Washington, D.C. 1990) offers a good overview of the field and its prospective importance. The term now denotes a broad spectrum of technologies and linked computer applications. Rather than eliminating or replacing human translators, machine translation technologies now facilitate their work, and serve other purposes ranging from aiding international product marketing to automating consumer assistance.

The worldwide information explosion which has increased the need for multi-lingual knowledge dissemination, and opportunities to service the international marketplace for information products. Translation technology is believed by some to be advantageous for international marketing and business transactions, and to be the kind of challenge for information technology developers that will produce beneficial side effects.

The fact that machine translation remains a significant challenge after thirty years should not hide the fact that progress has been made. Improvements in fundamental knowledge, and technical advances in the hardware and software of information technology, have contributed to the partial achievement of some of the intermediate goals of machine translation. There has been steady progress in relevant artificial intelligence software technology. Better indexing schemes, faster text searching algorithms, more clever ways to decompose and recompose language have been integrated to create new fields of knowledge processing, automated reasoning and inference. These have in turn become products in their own right as well as components in systems which facilitate the translation processes.

Computers, from the microprocessor to the supercomputer, have improved almost beyond belief, in raw processing power and in cost-performance because of advances in invisible, underlying technologies; the materials, devices, and tools by which machines are designed and built. Data storage and transmission devices have improved dramatically in density and throughput. Improved systems integration and design have contributed as much as any other factor to these developments. Recent results in parallel processing indicates that vastly more capable computers in every price and price-performance range are on the way. Such advances make previously difficult applications look easier, and machine translation is no different in this regard.

The historical lessons are straightforward. Significant advances in any complex computing application results from making headway across a broad range of underlying technologies. Making headway with the hard set of problems in machine translation means that a concerted but fairly broad approach is required.

ROLE OF NSF: PAST AND PRESENT ACTIVITIES

Recognizing this challenge, NSF has consistently supported basic research in machine translation and related technologies, within the context of broadening investment in computer and information science and engineering research.

For example, in FY 1990, CISE supported three projects in machine translation, at a total level of approximately \$350,000. These projects involve five institutions: New Mexico State University, Carnegie-Mellon University, New York University, Monmouth College, and City University of New York. We support a dozen projects in language generation, computational linguistics, and natural language understanding, at the total level of approximately \$1 million per year. These projects all deal with various aspects of the core technology critical to machine translation.

NSF has steadily developed its research programs in support of computing and information processing. These programs now cover all fields in computer and information science and engineering, with substantial investments in interdisciplinary research involving the relevant behavioral and social sciences. In addition to the underlying core technologies, we have supported research in a number of supporting technologies that directly or indirectly benefit the development or use of machine translation. These include research in:

Knowledge processing tools - algorithms, techniques and computer software to perform encoding, search, reasoning, planning and other symbol manipulating tasks associated with language processing or understanding.

Text or speech recognition - methods, architecture, computer algorithms designed to automatically capture language input, spoken or written, to a computer for improving the efficiency of information processing or the translation process.

Intelligent user interfaces - design of user interfaces to computers allowing more natural or better modalities of human-computer communications, including the use of pre- or post- editing facilities in machine translation.

The last area of research, intelligent user interfaces, is of particular relevance to computer-aided machine translation in that many applications of the translation process relies on the close interaction between the user and the machine. A three year award to Professor Alan Biermann of Duke University and Professor Robert Rodman of North Carolina State University, for example, is studying a new programming environment to allow the user interacting with a computer system in natural language by voice. This is comparable to major efforts in Japan, like the speech to speech Machine Translation project at the Advanced Telecommunications Research Institute (ATR).

INTERNATIONAL COOPERATION

NSF actively pursues international cooperation in basic research, and machine translation is no exception. The Japanese have been extremely active in the research and development of machine translation. There are a number of major R&D efforts in private industry, national laboratories, and universities. To assess the Japanese efforts first hand and to explore potential avenues of U.S. - Japan cooperation, NSF is co-sponsoring, with the Department of Commerce and the Defense Advanced Products Research Agency, an expert study team to visit major R&D sites in Japan. Their report will provide us with a close look at the state-of-the-art in Japan and an informed opinion of where the strategic opportunities are in advancing the progress and use of machine translation in the U.S.

COORDINATION WITH OTHER AGENCIES

NSF has been an active participant in federally coordinated strategic planning at all levels - from the FCCSET committee to the program level. Our efforts to coordinate activities with other federal agencies have been carried out in a broad range of computing activities - basic research, infrastructure development, and resources sharing. CISE has established several formal programs with DARPA in support of research in formal theory, computer vision, and speech understanding. New initiatives in language processing and machine translation are being explored at the present time.

CHALLENGES AND OPPORTUNITIES

We believe that the general directions posed by the NRC report described above are reasonable. First, we are increasing the efforts in multi-disciplinary basic research in the underlying technologies. By bringing together linguists, computer scientists and engineers in a cross disciplinary and integrated environment, we can expect more rapid progress in all aspects of the research: morphology, syntactic and semantic analysis, grammar development, knowledge representation, transfer mechanisms, and language generation.

Second, we are encouraging research aimed at better integrating the supporting technologies into machine translation. This means more group-sized experimental research in knowledge processing tools, speech/text recognition and intelligent user interfaces. Successful research in these areas will allow Machine Translation to be more effective as a technology for human enhancement. It can also contribute to stimulating the growth of a variety of other information technologies.

Third, and perhaps most important, we are developing the human resources and infrastructure for machine translation and related technologies. We are increasing the number of people and institutions capable of doing high quality research in all relevant areas of computing and information processing. At the same time we are building the infrastructure to support their efforts. Some practical steps include the development of sharable databases, bilingual texts and grammars and software for both research and education. One major step in this direction is underway in the deployment and improvement of NSFNET, which provides the pathway and opportunity for the interchange of information, computer-based tools, and more efficient coordination among researchers.

FEDERAL ROLE

In your invitation to testify you asked me to comment on what the Federal role should be, if any, in furthering research in machine translation technology. I believe that many federal agencies recognize the increasing importance of this technology to the nation's technology base and global interests. The agencies represented here are working together to assess the technology and to consider possible responses.

The general importance of information technologies is well understood. Under the umbrella of the Office of Science and Technology Policy, eight agencies have devised a strategic plan for improving a broad range of key computing and communications

technologies. This High Performance Computing program is presently wending its way through the process of budget construction for FY 1992, and several agencies have directed their FY 1991 requests toward its goals. Although not focused on machine translation, it will, if it comes to fruition, contribute substantially by creating a broad range of parallel computers, improving data storage and communication and technology, making a national research and education network available to the academic and industrial research communities, and alleviating the growing deficit in the appropriately skilled human resource base.

Improving international cooperation in the relevant areas of research and development is both desirable and essential. Efforts in developing comprehensive and up-to-date bilingual dictionaries and grammars, for example, are much better pursued by international teams. Users of this technology steeped in both the source and target languages could most effectively contribute to the development of standards and methods for system evaluation. Federal action in support of such international efforts is underway. Machine translation has been selected as an area of potential cooperation under the United States-Japan agreement for Cooperation in Science and Technology.

A task force, headed by the Commerce Department with the participation many agencies including NSF, has been working on ways to implement this joint activity. NSF is actively exploring avenues of joint research with the European Commission. Together with DARPA, we recently sponsored a series of workshops to study possibility cooperative research efforts in a variety of information technologies, including language processing. To make these efforts a success, we will have to change attitudes in many places about the support of international cooperation, and deal with mundane issues like adequate support of international travel and tougher problems of intellectual property rights and export control regulations.

Facilitating knowledge transfer and linkage to industry is also important. In the long run, such transfer and linkage is of course most effectively done through people, the trained scientists and engineers circulating among academic, government and industrial sectors.

Finally, and most critically, we must continue to make the investments in the kind of complex computer and information technologies represented in the broad area of machine translation. We have to sustain adequate, balanced long term investments in the full range of underlying fundamental science and engineering. Our private sector must efficiently use that knowledge to develop research products into commercially viable consumer products.

Mr. WALGREN. Okay, all right. Thank you, Dr. Brownstein.

I always wondered what went wrong with those monkeys and now I guess it's that they shifted over one frame on the keyboard, you know—to be or not to be, that is the—and then they must have shifted over. (Laughter) Oh, well.

Okay. And I gather, Dr. Chien, are there thoughts you'd like to add at this point?

Dr. CHIEN. Not at this point. But I'll be happy to help answer some questions later on.

Mr. WALGREN. Okay. Well, thank you very much.

Mr. Bostad?

Mr. BOSTAD. Thank you, Mr. Chairman. I would like to give some practical experience with the U.S. Air Force in working with machine translation well over 20 years, despite the vicissitudes in the reception of machine translation in this country. We got involved with the Russian-English system. It is still our bread-and-butter system; it has served us very well. Eighty percent of what we translate is Russian. We expect this to diminish perhaps a little bit as testing changes in the future.

FTD has played a unique role in the development of this system. We have been with one contractor in California for 20 years. For the time being we are in the process of writing another five-year contract.

The relationship works—has worked successfully because we have believed in machine translation and we've taken an active role in providing constant and intensive feedback to the contractor, and we've gained a great deal of knowledge about the system. So working hand-in-hand, we have succeeded with the Russian system and then with other systems.

We translate an immense amount of information, for information acquisition, in a very, very wide range of subject areas, from medicine to lasers, to you name it—and we translate it. This is a “general purpose—translate any type” of system.

Probably the most important thing that the system offers is—other than cost-effectiveness, which is documented—is the fast turnaround of information. We have researchers who have very stringent priorities, and it is our experience that we have completed many major projects by turning information around in a matter of weeks that would have taken a large cadre of human translators, in-house translators; or it would take an immense amount of time if we had it done commercially.

We believe at FTD that the most significant achievement that we have accomplished in recent years is opening up machine translation systems to non-translators. And this does not meet the criterion of extremely high quality automatic machine translation, but it is translation that is of a reasonably high level; it is quite usable; the dictionaries are extensive. And this is raw machine translation used by a non-translator. He simply—he's simply an analyst. A secretary calls up a menu, goes through a very simple process, types in a file—and the raw machine translation comes out in five seconds on the screen and is used for different types of information scanning, titles of books, paragraphs.

But what we found, somewhat to our amazement, was that complete short articles—5-6 pages—were being translated simply be-

cause it was faster than going through the bureaucracy of our department translation, which would take 10 days.

I believe that this is the way of the future. There is—this is showing increasing success at the European Commission in Luxembourg. It is being used by—and Carlsruhe—for information scanning.

When you have subject matter experts out there, and you have massive extensive dictionaries and you produce reasonable output, systems will be used.

In fact, we had—with some trepidation, at one point we put up the German system; had a minimal vocabulary of 50,000 words, and we had people already using what we considered an inferior system. In the meantime, we have now developed it to 150,000 words, and developed the system quite considerably.

So as has been stated before, once you make systems available, the need responds to that. We have people using the system who know that the translation division—Directorate—was already overtasked, and that we couldn't translate everything, so they would show some restraint in sending up translations. Now they can put in the random German translation that they would normally request, and they will get the information using raw machine translation.

The change in political climate and the constraints on resources, I believe, argue well for the future of machine translation.

From my point of view the government needs to support machine translation, and there are some language pairs that are of urgency at this time. If there was an Arabic-English system it would be used at this very moment.

There is a great deal of demand for a Chinese-English system because there are not enough competent translators out there.

I think there are a variety of systems that need to be developed in the future, and I think the government needs to sustain support.

That's all I have, Mr. Chairman.

[The prepared statement of Mr. Bostad follows:]

MACHINE TRANSLATION IN THE USAF

The purpose of this written statement is to describe the operational use of machine translation (MT) at the Foreign Technology Division (FTD), Wright-Patterson AFB, Ohio. The discussion includes a description of the various language pairs, rights of the systems, and their application. The statement makes some general observations about MT use in the government and suggests directions for further developments. Examples of Systran raw machine translations are appended to the written statement.

From the early 1960s with IBM's Mark I and Mark II systems, and from 1970 to the present with Systran, FTD has been actively involved in the development, use, and promotion of MT. FTD was one of the first large organizations to make MT operational on a day-to-day and year-to-year basis and prove the effectiveness of machine translation.

I. Machine Translation Systems

Russian-English The Russian system was the first system acquired and has undergone extensive development for twenty years. Linguistic development of major proportions ended three years ago, but there continues to be maintenance and minor modification of linguistic modules. Dictionary development is still extensive. The system now has in excess of 210,000 canonical single-word entries and 150,000 expressions. Very sophisticated dictionaries exist for avionics, computer science, medicine, physics, and chemistry. Within the last year-and-a-half there has been a major effort to develop naval, army, and economics dictionaries. Since 1974 FTD has played a significant role in the development of the system, having at one time as many as seven people working on dictionary development. Over the years approximately 50,000 pages of Russian text have been translated annually with the Russian-English system.

French-English, German-English These systems were acquired in 1983 and have undergone extensive development sponsored by FTD. The French system is widely used in Europe and much development now comes from Europe. The German system originally had 40,000 single-word entries. It now has approximately 150,000 entries, and with extensive development of software to decompose various types of word combinations the effective dictionary entries are in the range of 200,000. The French system is excellent; the German system, due to the nature of the language, is less sophisticated, but it is improving rapidly.

Other systems FTD and FBIS have jointly been developing the Systran Japanese-English System (JESS) since 1985. The system has been considerably enhanced and some testing has been carried out by FBIS for its use in translating titles of articles. The dictionary contains approximately 60,000 terms. FTD, in support of other government agencies, has acquired mainframe use of three

other systems: English-Spanish, English-Arabic, and Spanish-English. The first two systems have been extensively developed and are very good. The Spanish-English system is of tolerable quality for rough translations, but needs a couple of years of extensive development.

II. Rights of the Systems

The U.S. Air Force has rights to provide the Russian, French, and German systems free to other government agencies. Mainframe applications now in place are at FBIS, NTIC, FSTC, Nellis AFB, and NASA, Washington, D.C. (installation request). Other agencies, DoD components, and users gain access to on-line translations via a modem to FTD's mainframe computer. The English-Arabic system is leased, and resides on FTD's mainframe to support the U.S. 3rd Army and other Army components. The English-Spanish system is leased for the Inter-American Air Force Academy, Homestead AFB, Florida, but its use is made available to FTD. The Spanish-English system has been acquired on a nonlease basis for Air Force use and will shortly undergo extensive development.

III. Applications of MT

Partially-Edited Machine Translation Partially-edited MT is the standard product of the Directorate of Translations of FTD. The translations are used by scientific and technical researchers who need to stay abreast of foreign developments in a wide variety of technical fields. Most of the translations are produced from open-source literature, but there is also a substantial volume of classified documentations of foreign assets that is translated.

After some years of experimentation it was determined in 1974 that partially-edited machine translation was the only way to cope with the immense translation volume. Not only was it much faster than human translation, it was also much cheaper. There was also the awareness that much of the translation was done for information scanning, that perhaps only bits and pieces of a large translation contained relevant information, but this could not be determined prior to the actual translation.

In several surveys of the large FTD translation user population it was found that timeliness of the information was the most important concern of the end user. He would struggle with a somewhat inferior product if he could get the information quickly, especially since he had deadlines to meet and priority tasks to finish. We have had numerous occasions where we have had to complete high-priority translations on short notice - in a week or less. This could only be accomplished by MT. Machine translation has therefore succeeded in satisfying the requirements of the end user - fast access to the information content.

Machine translation has also proven much cheaper than either in-house or contracted human translation. We estimate that we would have to have an in-house cadre of 25 Russian translators to produce our yearly volume, or expend large sums on contractual translations. Our figures coincide with the cost-savings gains often cited in the literature: MT provides translations with savings of 40% as compared with the cheapest human translations.

Interactive Machine Translation Three years ago FTD embarked on what we consider one of the most important uses of MT: the use of raw machine translation for information scanning by nontranslators and -linguists. For over 15

years it was thought that the MT product was only useful if it was edited; we decided to free up the systems to anyone who wanted to make use of them in any way whatsoever. We wrote a user-friendly menu and made the system available on 1400 PCs within FTD. The analyst, intelligence aide, or secretary types in a file, sends it to be translated, and the translation appears on the display in five to 20 seconds, depending on the length of the file. The original intent of this application of MT - what we call interactive MT - was to translate titles of books, tables of contents, figure captions, and short paragraphs. But in reality the use has extended to the translation of complete short articles. The frequency of use of this application has far exceeded all expectations. The system is now accessed by some 200 users across FTD with average usage per month as follows: Russian - 200 accesses/month, German - 30 accesses/month, and French - 25 accesses/month.

This application of MT is of increasing importance. The Commission of the European Communities in Luxembourg, which has 13 Systran language pairs, is also achieving notable successes with the use of MT by nontranslators. Bilingual secretaries use it, it is used to quickly translate internal documents for information prior to meetings, and it is used to translate newspaper articles to keep up with fast-breaking current affairs. Again, the translations do not have to be perfect, but there are demands on the MT systems in order for raw MT to be useful. The systems must have robust algorithms that can handle a variety of sentence structures and styles and, most importantly, they must feature massive, comprehensive dictionaries in a wide range of subject fields. If the systems produce excessively poor translations, or not enough words are found, they will not be used.

MT within the Government The most important application of MT within the government is for the translation of large volumes of information for assimilation, i.e., to find out what is being said. This is FTD's use, but it is also the primary use of many DoD and intelligence components. By extension, "translation for assimilation" can include the translation of patents, abstracts from foreign databases, for immigration documentation, etc. There is also a specific need, albeit less, for the translation of what is called "information for dissemination." Examples are the Inter-American Air Force Academy which translates English manuals into Spanish, the School of the Americas with a similar requirement, and whatever documentation must accompany equipment sold by the government, e.g., the sale of military aircraft to a foreign country, where the documentation preferably would be in the language of the buyer country instead of English. This use of MT more closely parallels commercial applications of MT where great accuracy is required and very specific sublanguages (technical vocabularies) are used. However, it would seem that the use of MT for information assimilation will remain the primary application within the government.

III. The Future of Machine Translation

The future of MT is secure and it is here to stay. There are now many well-documented instances of successful and cost-effective applications. In applications of both dissemination and assimilation of information significant increases in translation productivity have been recorded. Of primary importance is the understanding that there are a plethora of user requirements and that MT can satisfy many of them, and that the ultimate test is whether the users' needs are met. Too often in the government, standards for translation quality have been set by translation shops that reflect the perfectionist standards of human

translators. In times of stringent budget constraints when resources are limited, more and more translation organizations which previously defined quality from their own point of view are now willing to ask the user what he really wants and what quality of translation he is willing to accept. In many cases it has been found that MT is perfectly acceptable.

There are now at hand many tools that make MT much more efficient and useful. The most important development is the miniaturization of hardware that has allowed sophisticated systems to run on microcomputers and PCs. Systran, for example, has been downloaded to run on stand-alone IBM PCs by FTD. It has successfully run on an IBM PC AT/370 and a PS-2 7437 workstation. There have also been impressive developments in pre-editing software that detects MT ambiguities, the use of controlled source languages, post-editing software for rapid post-editing, rapid dictionary-update software, powerful word-processing packages that allow global replacement of words, and the incorporation of spelling checkers into the translation process.

There are several vendors, including LOCOS, METAL (Siemens), and Systran that have a proven track record of developing successful operational systems. The government has access, through buying or leasing, to a wide range of language pairs. But some languages require development and the need has become urgent with the changing political climate. These languages are Arabic, Chinese, Farsi, and Korean that will translate into English. The investment costs will be many millions of dollars, but usable systems could be operational in a few years. There are also good reasons for further research in MT. However, I believe that government-supported research, at this stage of the game, should be much more selective than it was in the early and middle sixties. A pragmatic approach, I believe, is the example of the Commission of the European Communities. They embarked on the EUROTRA project, but never gave up on parallel development of the many Systran systems they have acquired. A combination of research and pragmatic development, I believe, is the best approach.

Public opinion has now shifted in favor of MT. It is written up in the popular press and more people are being exposed to it, the notable example being MT on the Minitel in France. It has found a niche as one of the tools for coping with the information explosion, and its use will continue to grow.

\$MH .EBR88047301 \$MHX \$L1 \$SC214D

\$L2 .Page 1.

\$L3 .TITLE RASPROSTRANENIE I REGULIROVANIE DIFRAKSIONNYX UKV POLEI. .ENDTITLE.

\$L1 N. B. Cimitdorjiev, P. N. Dagurov, H. L. Lomuxin.

\$NP .Page 2.

\$L2 \$T V monografii privod4ts4 rezul'taty issledovani4 prostranstvenno-vremennoi struktury UKV pol4 pri mnogolucevom difrakcionnom rasprostraneni4 santi- i deqimetrovyx voln. Izlagahts4 metodu uvelicheni4 urovn4 signala i umen6veni4 vli4ni4 mewah5ego izluceni4 na prizemnyx peresecannyx trassax s ispol6zovaniem difraktornyx 3kranov. Rassmatrivahts4 sposoby uvelicheni4 razv4zki antenn, raspolojennyx vblizi provod4s4x tel. \$L3 \$T D14 radiofizikov, spetsialistov po prikladnoi 3lektrodinamike i antennoi tekhnike.

\$NP .Page 3.

\$L3 .TITLE PREDISLOVIE. .ENDTITLE.

\$L3 \$T Pri rasprostraneni4 UKV v peresecennoi mestnosti osnovnym mexanizmom peredaci 3nergii v oblast6 teni na trassax nebol6wol i srednel i srednel prot4jennosti 4vl4ets4 difrakci4 na dominiruh5ix prep4tstvii4x rel6efa. Vo mnogix sluca4x uroven6 difrakcionnogo pol4 mojet byt6 dostatocno vysokim dl4 sozdani4 linii zagorizontnoi sv4zi.

\$L3 \$T S drugoi storony, neobxodimost6 funkcionirovani4 radiosistem razlichnogo naznacheni4 v uslovi4x obespeчени4 3lektromagnitnoi sovместимости (3MS) trebuet znani4 osobennostel prostranstvenno-vremennoi struktury difrakcionnyx polei.

\$L3 \$T V diapazone santi- i deqimetrovyx voln pole na zakrytyx trassax, kak pravilo, imeet prostranstvenno-neodnorodnuh strukturu, obuslovlennuh mnogolucevost6h signala vsledstvie difrakcii i rasse4ni4 na nerovnost4x grebnel 3kraniruh5ix prep4tstvi4. Na real6nyx prizemnyx trassax prostranstvenno-vremenna4 neodnorodnost6 pokazatel4 prelomleni4 troposfery, vyzyvah5a4 po4vlenie turbulentnyx i sloistyx neodnorodnostel, izmenenie uslovi4 refrakcii, privodit k deformaqii prostranstvennoi struktury pol4 i vremennym fluktuaci4m signala. Zamirani4 signala pri 3tom doljny zaviset6 ot osobennostel tonkol prostranstvennoi struktury pol4. Odnako obosnovannye metody i neobxodimye dl4 praktičeskix rascetov dannye, pozvol4h5ie opredelit6 uroven6 signala i ego fluktuacii v kajdom konkretnom slucae, poka otsutstvuh. Po3tomu bol6woe znachenie imeht 3ksperimental6nye issledovani4 mnogolucevogo difrakcionnogo rasprostraneni4 pol4 v real6nyx uslovi4x, 4vl4h5ies4 osnovoi dl4 razrabotki metodov rasceta trass s podobnym mexanizmom rasprostraneni4.

\$L3 \$T V tex sluca4x, kogda parametry difrakcionnogo signala ne udovletvor4ht uslovi4m normal6noi raboty vzaimodelstvuh5ix radiosredstv, po4vl4ets4 neobxodimost6 regulirovani4 prizemnogo pol4. Odnim iz 3ffektivnyx metodov reweni4 3tol zadaci 4vl4ets4 primenenie difrakcionnyx radioneprozrachnyx 3kranov.

\$L2 .END .SECTION.

08-27-90
MICROFICHE HEADER
EBR88047301
08-27-90

PAGE 1

PAGE 1

SUBJ CODE 214D

MT VERSION- PROG 16 JUL 1990 DICT

Page 1.

PROPAGATION AND CONTROL OF DIFFRACTION VHF FIELDS.

N. B. Chimitdorzhiev, P. M. Dagurov, Yu. L. Lomukhin.

08-27-90

PAGE 2

Page 2.

In monograph results of investigation of space-time structure of VHF field during multiple-pronged diffraction propagation of centi- and decimeter waves are given. Are presented to method increases in the level of signal and decrease of the effect of background radiation on the surface crossed routes with the use of diffraction shields. The methods of an increase in decoupling the antennas, located near conductive bodies, are examined.

For radiophysicists, specialists in applied electrodynamics and antenna to technology.

08-27-90

PAGE 3

Page 3.

PREFACE.

During propagation of VHF in broken ground diffraction on prevailing obstructions of relief is fundamental mechanism of energy transfer into shadow zone on routes of small and average and average extent. In many instances the level of diffraction field can be sufficiently high for compiling of the lines of communications below the horizon.

On the other hand, need for functioning of radio systems of different designation/purpose under conditions of guaranteeing electromagnetic compatibility (EMS) requires knowledge of special features of space-time structure of diffraction fields.

In the range of centi- and decimeter waves field on closed routes, as a rule, has space-inhomogeneous structure, caused by multi-beam character of signal as a result of diffraction and scattering on inequalities of peaks of shielding obstructions. On the real surface routes the space-time heterogeneity of the refractive index of the troposphere, which causes the appearance of turbulent and laminar heterogeneities, a change in the conditions of refraction, lead to the strain of the three-dimensional structure of field and the temporary/time fluctuations of signal. Signal fading in this

08-27-90

PAGE 4

case must depend on the special features of the thin three-dimensional structure of field. However, the substantiated methods and necessary for the practical calculations given, which make it possible to determine the level of signal and its fluctuation in each specific case when they are absent. Therefore the experimental studies of the multiple-pronged diffraction propagation of field in the actual conditions, which are base for developing the methods of calculation of routes with a similar mechanism of propagation, have high value.

When parameters of diffraction signal do not satisfy conditions for normal operation of interacting radio aids, appears need of controlling surface field. One of the effective methods of the solution of this problem is the application of diffraction radio-opaque shields.

END SECTION.

\$MH .EBR87102500 \$MHX

\$L1 \$SC3D

\$L2 .Page 62.

\$L3 .TITLE Programmiersprachen. .ENDTITLE.

\$L1 Von Lawrence G. Tesler.

\$L2 \$T Mit dem Computer in dessen Sprache zu verkehren ist sehr muhsam. Deshalb wurden Programmiersprachen entwickelt, die es erlauben, mit dem Rechner in menschengemaesserer Weise zu kommunizieren. Die modernsten verzichten ganz auf Handlungsanweisungen und ueberlassen es dem Computer zu entscheiden, wie er ein Problem anhand der eingegebenen Fakten loest.

\$L3 \$T Eine Programmiersprache ist mehr als eine Notation, um einem Computer Anweisungen zu geben. Eine Sprache und die Software, die sie "versteht", koennen den Computer voellig umgestalten, aus ihm eine ganz andere Maschine machen.

\$L3 \$T Die Hardware eines typischen Computers besteht aus Registern, Speicherzellen, Addierern und so weiter. Wenn ein Programmierer die Maschinensprache des Computers benutzt, sind dies die Elemente, die er im Kopf haben muss. Eine neue Sprache aber bringt auch ein neues Modell der Maschine mit sich. Obwohl die Hardware unveraendert bfeibt, kann der Programmierer in anderen Kategorien denken: in Variablen anstatt in Speicherzellen, in Dateien anstatt in Einund Ausgabekanaelen und in algebraischen Formeln anstatt in Registern und Addierern. Einige Sprachen verleihen dem Computer sogar gleichsam eine gesplaltene Persoenlichkeit: Er wird zu einem Verbund unabhaengiger Stellen, die ihre eigenen Berechnungen vornehmen und untereinander kommunizieren.

\$L3 \$T Die Zahl der Programmiersprachen und ihrer Dialekte geht in die Hunderte, wenn nicht gar Tausende. Zwar mag die Zahl der natuerlichen Sprachen noch hoeher liegen, aber in gewisser Hinsicht sind Programmiersprachen vielfaeltiger als das menschliche Sprachspektrum. Jede Programmiersprache hat ja nicht

nur ihre eigene, charakteristische Grammatik und Syntax, sondern auch ihre eigene Art, Ideen auszudruecken.

\$L3 \$T Im Prinzip waere es mit jeder Programmiersprache moeglich, die meisten Aufgaben fuer einen computer zu formulieren; doch die Programme saehen sehr unterschiedlich aus. Ausserdem liesse sich eine gegebene Aufgabe mit einigen Sprachen leichter programmieren als mit anderen. Ich werde hier einige der gebräuchlichsten Programmiersprachen beschreiben und versuchen, die Gemeinsamkeiten und Unterschiede zwischen ihnen deutlich zu machen.

\$L3 .TITLE Grundprinzipien der Programmentwicklung. .ENDTITLE.

\$L3 \$T Betrachten wir zuhaechst ein kurzes Programm in LOGO, einer Sprache, die in den spaeten sechziger Jahren von Seymour Papert und seinen Mitarbeitern am Massachusetts Institute of Technology (MIT) erfunden wurde. Eine interessante Eigenschaft von LOGO ist, dass sich damit eine "Schildkroete" steuern laesst - eine Art Miniroboter, der sich vorwaerts und rueckwaerts bewegen, sich auf der Stelle drehen und einen stift heben und senken kann, der seinen Weg mitzeichnet. Oft ist die Schildkroete kein wirkliches Geraet, sondern wird nur auf einem Bildschirm simuliert.

\$L3 \$T Die erste Version des Programms lautet \$L8. Sie besteht gaenzlich aus Befehlen an die Schildkroete. Nach der Anweisung, den Stift zu senken, folgen fuefmal hintereinander die zwei Befehle forward 50 ("vorwaerts 50") und right 144 ("rechts 144"); danach wird der Stift wieder angehoben. Wenn die Schildkroete den Anweisungen folgt, zeichnet sie einen fuefzackigen Stern. Das Kommando forward 50 naemlich bewirkt, dass sie eine 50 Einheiten lange gerade Linie erzeugt; right 144 bedeutet eine Drehung um 144 Grad im Uhrzeigersinn; um diesen Betrag aendert die Schildkroete an jeder Spitze des fuefzackigen Sterns jeweils ihre Laufrichtung.

\$L3 \$T Koennte man einem Computer seine Absichten nur dadurch mitteilen, dass man alle Anweisungen einzeln hintereinander auflistet, waere es um die Programmierkunst schlecht bestellt. Tatsaechlich gewaehren LOGO und andere Programmiersprachen eine Reihe von Moeglichkeiten, Anweisungen einfacher und allgemeingueeltiger zu gestalten. In diesem Fall ist es die fueffache Wiederholung der forward- und right-Anweisung, die offensichtlich einer Verbesserung bedarf. Wo immer moeglich,

vermeidet ein Programmierer, etwas mehr als einmal zu schreiben - und zwar nicht nur, weil das Tippen laestig ist. Je kleiner naemlich ein Programm, desto weniger Platz braucht es im Speicher. Ausserdem erhoehrt das Wiederholen die Wahrscheinlichkeit von Tippfehlern, besonders wenn das Programm ueberarbeitet wird. Die Wiederholung laesst sich vermeiden, wenn man die fuenf Kommandos zum Bewegen der Schildkroete durch die Anweisung repeat 5 [forward 50 right 144] ersetzt (repeat ist das englische Wort fuer "wiederhole").

\$L3 \$T Angenommen, der Programmierer moechte nun einen neunzackigen Stern mit 80 Einheiten langen Kanten zeichnen. Das waere durch eine Anweisung wie repeat 9 [forward 80 right 160] zu erreichen.- Diesem Befehl liegt offensichtlich die gleiche Programmstruktur zugrunde wie dem vorhergehenden; Unterschiede bestehen nur im Detail.

EBR07102500

SUBJ CODE 3D

MT VERSION- PROG DICT

Page 62.

Programming languages.

Of Lawrence g. Tesler.

With the computer in its language to operate is very laborious. Therefore programming languages were developed, which permit it communicate with the processor in menschengemaesserer way. The most modern do completely without action statements and leave it to the computer to decide, how it solves a problem on the basis the input facts.

A programming language is more than a notation, in order to give to a computer statements. A language and the software, which "understand, can" it the computer completely transform, from him one other machine make completely.

The hardware of a typical computer continues to consist of registers, memory cells, adders and so. If a programmer uses the machine language of the computer, this is the items, which it must have in the heading. A new language however also brings a new model the machine with itself. Although the hardware remains unchanged, the programmer in other categories can think: in variables instead of in memory cells, in files instead of in Einund output channels and in algebraic formulas instead of in registers and adders. Some languages even lend as it were a split personality to the computer: it becomes a network of independent places, which communicate their own calculations distinguished and among themselves.

The number of the programming languages and their dialects goes into the hundreds, if not even thousands. The number of the natural languages may still be situated more highly, but in certain regard programming languages are more varied than the human Sprachspektrum. Each programming language has not only their own, characteristic grammar and syntax, but also their own type to express ideas.

In principle it would be possible with each programming language to formulate the most functions for a computer; the

programs would look nevertheless very differently. Additionally a given function with some languages could more easily be programmed than with others. I will describe here some the most common programming languages and will try to make the thing in common and differences between them clear.

Basic principles of the program development.

We regard first a short program in LOGO, a language, which was invented in the late sixties by Seymour Papert and its employees at the Massachusetts Institute of Technology (along). An interesting characteristic of LOGO is mitzeichnet that thereby one "can be controlled Schildkroete", - a type mini robot, which moves forward and backwards, on the place to turn and a pin lift and lower can, its path. Often Schildkroete is not a real device, but on a display only is simulated.

The first version of the program reads

It completely consists of instruction to Schildkroete. After the statement to lower the pin consecutively the two instruction forward 50 follow five times ("forward of 50") and right 144 ("on the right of 144"); after it is raised the pin again. If Schildkroete follows the statements, she draws fuenfzackigen asterisk. The command forward 50 causes that it means 50 units a long even line; right 144 a turn by 144 degrees in the clockwise direction; around this amount modifies Schildkroete in each point fuenfzackigen asterisk in each case their direction of travel produced.

If one could only indicate its intentions to a computer by the fact the fact that one lists all statements consecutively individually would be it with Programmierkunst badly. LOGO and other programming languages actually grant a set of possibilities of arranging statements simpler and more generally accepted. In this case it is the fivefold repetition of the forward and right-Anweisung, obviously an enhancement requires. Where always possible, a programmer does not avoid to write somewhat more than once - only, because tapping is annoying. The smaller a program, workstation all the fewer needs it in the memory. Additionally repeating increases the probability by typing errors, particularly if the program is revised. The repetition can be avoided, if one

repeat the five commands for moving Schildkroete by the statement to 5 [forward 50 right 144] replaced (the English word for "is repeat repeats").

Assumed now, the programmer would like to draw neunzackigen asterisk with 80 units long edges. That would be by a statement as repeat 9 [forward 80 right 160] to achieve. Obviously the same program structure is appropriate for this instruction as the base as preceding; differences only exists in the detail.

\$MH .EBR87F005300 \$MHX

\$L1 \$SC314D

\$L2 .Page 20.

\$L3 .TITLE La prochaine r3evolution informatique. .ENDTITLE.

\$L3 \$T En moins d'un demi-si2ecle, les ordinateurs sont devenus instruments indispensables dans l'industrie; les futurs ordinateurs seront moins dix fois plus performants et fourniront une aide intellectuelle dans tous les domaines. .ENDTITLE.

\$L1 par Abraham Peled.

\$L3 \$T Quand il apparut, l'ordinateur n'3etait qu'une curiosit3e de laboratoire; c'est aujourd'hui un 3el3ement vital de notre soci3et3e. Les activit3es financi2eres, industrielles et de transport ne pourraient plus fonctionner sans les traitements et les 3echanges 3electroniques; l'ordinateur participe aujourd'hui 2a la conception de mat3eriaux et de produits pharmaceutiques, et aux recherches de physique fondamentale sur la nature de l'3energie et de la mati2ere. L'activit3e 3economique et intellectuelle conna4it gr4ace 2a cet instrument irrempla3eable un d3veloppement sans pr3ec3edent, et ceux qui participent 2a la croissance (exponentielle!) du monde informatique vivent sans doute un des moments les plus exaltants de l'histoire des sciences et des techniques.

\$L3 \$T Un nouvel 3episode de l'histoire de l'informatique d3bute aujourd'hui: dans les dix ann3ees 2a venir appara4itront de nouveaux ordinateurs dix fois plus puissants, plus 3elabor3es et plus maniables que ceux que nous utilisons. Gr4ace aux extraordinaires progr2es des logiciels et des mat3eriels, l'ordinateur va devenir un outil intellectuel aussi indispensable et courant que le t3el3ephone. Des dispositifs plus "naturels" de communication - notamment de nouveaux types d'affichages - entre l'Homme et l'ordinateur permettront 2a tous d'utiliser cet outil; de vastes r3eseaux relieront tous les utilisateurs de l'informatique: les m3edecins 2a la recherche de diagnostics difficiles, les financiers pr3eparant un plan d'investissement, les 3aerodynamiciens concevant des avions, les 3etudiants pr3eparant leurs examens, etc. b3en3eficieront tous des fonctions

modernes des fonctions modernes des réseaux.

\$L3 \$T La révolution informatique qui se prépare résulte de notre profond changement de mentalité face à l'ordinateur, mais elle est sous-tendue par les transformations subies par le monde industriel depuis la fin de la Seconde Guerre mondiale. La miniaturisation des composants électroniques a jeté l'élément déterminant de cet essor: grâce à elle, le coût d'un calcul informatique diminue de 20 à 30 pour cent par an depuis 30 ans. En effet la taille d'un transistor, d'une porte logique ou de tout élément de circuit diminue, la vitesse des opérations effectuées augmente en proportion inverse et la densité des éléments placés sur un composant croît de façon exponentielle.

\$L3 .TITLE La miniaturisation des composants électroniques.
.ENDTITLE.

\$L3 \$T La miniaturisation des composants se poursuivra probablement au rythme actuel pendant environ 10 ou 15 ans grâce à l'amélioration - difficile, mais possible - des techniques actuellement mises en œuvre (voir Les puces de l'informatique de demain, par James Meindl, Page 42). Avec la lithographie par rayons X, utilisant le rayonnement synchrotron, avec de nouveaux matériaux et de nouvelles conceptions de circuits intégrés, la densité des composants sera sans doute multipliée de 20 à 40 fois, de sorte que les ordinateurs de demain seront probablement 6 à 12 fois plus rapides que ceux d'aujourd'hui.

\$L3 \$T Les progrès résulteront également de l'introduction du parallélisme dans les ordinateurs: contrairement aux ordinateurs classiques, qui effectuent les calculs les uns après les autres, les ordinateurs parallèles effectuent simultanément plusieurs - voire toutes - les opérations d'un problème et sont par conséquent beaucoup plus rapides (voir L'architecture des futurs ordinateurs, par Geoffrey Fox et Paul Messina, page 30).

\$L3 \$T Si le rythme du progrès technique dépend de la miniaturisation, l'intégration de l'informatique dans la société est tributaire des réalisations et des commercialisations des logiciels: ce sont les programmes qui transforment les potentialités des ordinateurs en réalisations adaptées aux problèmes de l'utilisateur. Il n'y a pas de voie

royale pour l'amélioration des logiciels, et les progrès proviendront à la fois d'études sur les structures, sur les langages et sur les environnements de programmation (voir Les nouvelles méthodes de programmation, par David Gelernter, page 52).

\$L3 \$T La structure d'un programme est la façon dont sont agencées en sousprogrammes les opérations élémentaires, et les sous-programmes en modules qui, ensemble, résolvent les problèmes posés. Les langages sont les ensembles d'instructions commandant les opérations; plus les langages sont "de haut niveau", moins l'utilisateur doit préciser la marche à suivre pour résoudre le problème traité. Aujourd'hui, avec les langages les plus évolués, l'utilisateur peut se contenter de taper la description mathématique ou logique du problème qu'il souhaite résoudre, et l'ordinateur détermine seul le détail des opérations.

\$L3 \$T L'environnement de programmation, c'est-à-dire l'ensemble de moyens physiques et logiques dont dispose l'utilisateur pour transmettre les instructions, a beaucoup évolué; on manipulait jadis des interrupteurs - on utilise aujourd'hui des claviers et des souris; on remplissait autrefois des organigrammes et des formulaires de programmation complexes - on dispose aujourd'hui de menus interactifs et de représentations graphiques.

KBR67F05300

SUBJ CODE 314D

MT VERSION- PROG

DICT

Page 20.

Next data-processing revolution.

Less than one half-century, the computers became essential instruments in industry; the future computers will be less ten time the most powerful and will provide a intellectual assistance in all the fields.

by Abraham Peled.

When it appeared, the computer was only one curiosity of laboratory; it is a vital element of our company today. The financial, industrial activities and of transport could not function any more without the electronic processing and exchanges; the computer takes part today in the material design and of medicinal products, and in the searches of physics fundamental on the nature of energy and matter. The economic and intellectual activity knows thanks to this irreplaceable instrument a development without precedent, and those which take part in the growth (exponential !) of the data-processing world live one of the moments more exaltants of the history of sciences and technology undoubtedly.

A new episode of the history of data processing begins today: in the next the ten years will appear new computers ten time more powerful, more elaborate and more handy than those than we use. Thanks to extraordinary progress of the software and the hardwares, the computer will become such a essential intellectual tool and current that the telephone. "Natural" devices more of communication - in particular of new types of displays - between the Man and the computer will allow all using this tool; of vast networks will connect all the users of data processing: the doctors in the search of difficult diagnoses, the financiers preparing a plan of investment, the aerodynamicists designing aircraft, the students preparing their examinations, etc. will profit all by the modern functions of the modern functions of the networks.

The data-processing revolution which prepares results from our deep change of mentality vis-a-vis the computer, but it is underlain

by the transformations undergone by the industrial world since the end of the Second world War. The miniaturization of the electronic components was the element determining of this rise: thanks to it, the cost of a data-processing calculation has decreased per annum by 20 to 30 percent for 30 years. Indeed the size of a transistor, a logical door or of any element of circuit decreases, the speed of the operations carried out increases in opposite proportion and the density of the elements placed on a component grows in an exponential way.

Miniaturization of the electronic components.

The miniaturization of the components will probably continue at the current rhythm during approximately 10 or 15 years thanks to the improvement - difficult, but possible - of techniques currently implemented (see The chips of data processing of tomorrow, by James Meindl, Page 42). With the lithography by x-rays, using the synchrotron radiation, with new materials and new designs of integrated circuits, the density of the components will undoubtedly be multiplied from 20 to 40 times, so that the computers of tomorrow will be probably 6 to 12 time the fastest than those of today.

Progress will also result from the introduction of parallelism into the computers: contrary to the conventional computers, which carry out calculations the ones after the others, the parallel computers carry out simultaneously the several - even all - operations of a problem and are consequently much faster (see Architecture of the future computers, by Geoffrey Fox and Paul Messina, page 30).

If the rhythm of technical progress depends on the miniaturization, the integration of data processing in the company is tributary of the achievements and marketings of the software: in fact the programs transform the potentialities of the computers into achievements adapted to the problems of the user. There is no royal channel for the improvement of the software, and progress will come at the same time from studies on the structures, the languages and on the environments of programming (see The new methods of programming, by David Gelernter, page 52).

The structure of a program is the way whose is arranged in subroutines the elementary operations, and the subroutines in modules which, together, solve the problems arising. The languages

are the assemblies of instructions controlling the operations; more the languages are "high level", less the user must specify the procedure to solve the dealt with problem. Today, with the the most evolved languages, the user can restrict oneself typing the mathematical or logical description of the problem which he wishes to solve, and the computer determines only the detail of the operations.

The environment of programming, i.e. the whole of average the physics and logics available to the user to transmit the instructions, evolved much; formerly switches were handled - keyboards and mice are used today; flow charts and complex forms with programming were filled formerly - one has today of interactive fragments and charts.

The Origins of Indo-European Languages

Almost all European languages are members of a single family. The author contends that they spread not by conquest, as has been thought, but along with the peaceful diffusion of agriculture

by Colin Renfrew

One of the most contentious problems in the entire field of archaeology and prehistory is the explanation of the remarkable relations that link nearly all the European languages, many of the languages spoken in India and Pakistan and some of those in the lands between. For more than two centuries it has been known that the "Indo-European" languages are related. But what prehistoric process underlies that relationship? How did related languages come to be spoken over such a wide area? What implications does their distribution hold for European prehistory and history? (Keep in mind as well that as a result of the colonial expansion of the 16th to 19th centuries, more people speak Indo-European languages than speak those of any other language group.)

The traditional view of the spread of the Indo-European languages holds that an Ur-language, ancestor to all the others, was spoken by nomadic horsemen who lived in what is now western Russia north of the Black Sea near the beginning of the Bronze Age. As these mounted warriors roamed over greater and greater expanses, they conquered the indigenous peoples and imposed their own proto-Indo-Euro-

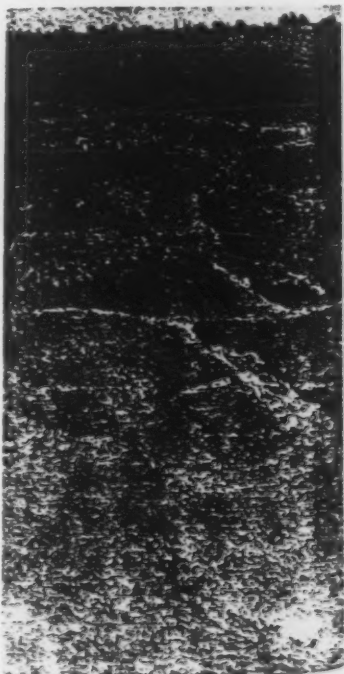
pean language, which in the course of succeeding centuries evolved in local areas into the European languages we know today.

In recent years, however, many scholars, particularly archaeologists, have become dissatisfied with the traditional explanation. I have analyzed the arguments for the accepted view and found them unconvincing. In this article I offer a different view, based on new insights into how cultural change comes about. According to this view, the spread of the Indo-European languages did not require conquest. On the contrary, it was likely to have been a peaceful diffusion linked to the spread of agriculture from its origins in Anatolia and the Near East. This proposed solution, which differs markedly from the accepted one, has profound implications for European prehistory and for linguistic studies of Indo-European languages.

The starting point of the problem of the origins of Indo-European is not archaeological but linguistic. When linguists look at the languages of Europe, they quickly perceive that these languages are related. The connections can be seen in vocabulary, grammar and phonology (rules for pronunciation). To illustrate the relatedness in vocabulary, it is sufficient to compare the words for the numbers from one to 10 in several Indo-European languages [see illustration on page 109]. Such a comparison makes it clear that there are significant similarities among many European languages and also Sanskrit, the language of the earliest literary texts of India, but that languages such as Chinese or Japanese are not members of the same family.

More detailed comparisons enable

of Europe further, into families. The earliest family to be recognized, that of the Romance languages—all known to be descended from Latin—includes French, Italian, Spanish, Portuguese and Romanian. The Slavonic languages include Russian, Polish, Czech, Slovak,



CATAL HUYUK is an early agricultural site in the central Anatolian plain in

COLIN RENFREW is Disney Professor of Archaeology at the University of Cambridge and Master of Jesus College there. He has conducted archaeological excavations in Greece and in the British Isles. His previous articles for *Scientific American* were on radiocarbon dating, the obsidian trade and the megalithic monuments of Europe. The arguments made here are presented in greater detail in Renfrew's book, *Before Farming and Language: The Origins of Indo-European*, published by Cambridge University Press in 1987.

Serbo-Croat and Bulgarian. The Germanic languages include German, Norwegian, Danish and Swedish. These families, which link the most closely related tongues, can themselves be regrouped to form the larger Indo-European language family. Indeed, only a few European languages (Hungarian, Finnish and Basque, for example) are not members of this family.

How did this complex arrangement arise? The Romance languages served as the first model for answering the question. Even to someone with no knowledge of Latin, the profound similarities among Romance languages would have made it natural to suggest that they were derived from a common ancestor. On the assumption that the shared characteristics of these languages came from the common progenitor (whereas the divergences arose later, as the languages diverged), it would have been possible to reconstruct many of the characteristics of the original proto-language. In much the same way it became clear that the

branches of the Indo-European family could be studied and a hypothetical family tree constructed, reaching back to a common ancestor: proto-Indo-European.

This is the tree approach, which was pioneered in the early 1860's by the German philologist August Schleicher. It provides the framework for the way most historical linguists still think about the development of language families. The basic process represented by the tree model is one of divergence: when languages become isolated from one other, they differ increasingly, and dialects gradually differentiate until they become separate languages.

Divergence is by no means the only possible tendency in language evolution. Only a decade after Schleicher proposed his tree hypothesis, another German linguist, Johannes Schmidt, introduced a "wave" model in which linguistic changes spread like waves, leading ultimately to convergence—that is, growing similarity among lan-

guages that were initially quite different. In 1939 the Soviet linguist N. S. Trubetskoy went so far as to suggest that the resemblances among the Indo-European languages might have come about this way. Today, however, most linguists dismiss Trubetskoy's suggestion and think primarily in terms of linguistic family trees.

Although the starting point of the problem of Indo-European origins may be linguistic, its solution certainly entails archaeology, which offers a means of testing the linguistic hypotheses. The archaeologists began tackling the problem in the early 1900's. At that time the archaeology of the Romans, Greeks and Celts was becoming increasingly well understood back to the early historical period, perhaps as far back as the first millennium B.C. Moreover, archaeological finds were being made that promised to extend scholarly understanding deep into prehistory, indeed back to Paleolithic (Old Stone



modern Turkey. Farming arose here and at nearby sites by about 7000 B.C.; not long afterward it began to spread north

into Europe. The author contends that the prototypical form of the Indo-European language spread along with agriculture.

SYSTRAN English-Spanish Translation
 SYSTRAN Translation Systems, Inc.
 La Jolla, Ca. 4/13/90

MICROFICHE HEADER

EBR90jaa

SUBJ CODE 124D

MT VERSION- PROG

DICT

Los orígenes de los lenguajes indoeuropeos.

Colin Renfrew.

Uno de los problemas más discutibles en el campo entero del archaeology y de la prehistoria es la explicación de las relaciones notables que ligan casi todos los lenguajes europeos, de muchas de los lenguajes hablados en la India y Paquistán y de algo de esas en las tierras en medio. Pero qué proceso prehistórico ser la base de esa relación? Cómo los lenguajes relacionados vinieron ser hablados concluido un área tan amplia? Qué implicaciones su distribución celebra para la prehistoria y la historia europeas? (Tenga presente también que como resultado de la extensión colonial de los décimosextos a diecinueveavo siglos, más gente habla los lenguajes indoeuropeos que hablan esos de cualquier otro grupo del lenguaje).

La opinión tradicional de la extensión de los lenguajes indoeuropeos sostener que un Ur-lenguaje, antepasado a todos los otros, fue hablado por los horsemen nómadas que vivieron en cuál ahora está norte occidental de Rusia del Mar Negro cerca del principio de la edad de bronce. Mientras que estos guerreros montados vagaron concluido extensiones más mayores y más mayores, conquistaron la gente indígena e impusieron su propio lenguaje proto-Indo-Europeo, que en el curso de los siglos que tienen éxito se desarrolló en las áreas locales en los lenguajes europeos que nosotros sabemos hoy.

En los años recientes, sin embargo, muchos eruditos, particularmente arqueólogos, se han descontentado con la explicación tradicional. Yo he analizado las discusiones para la opinión aceptada y las encontré poco convincentes. En este artículo yo

ofrezco una diversa opinión, basada en las nuevas penetraciones en cómo viene alrededor el cambio cultural. Según esta opinión, la extensión de los lenguajes indoeuropeos no requirió la conquista. En el contrario, era probable ser una difusión pacífica ligada a la extensión de la agricultura de sus orígenes en Anatolia y el este cercano. Esta solución propuesta, que diferencia marcado aceptada, tiene implicaciones profundas para la prehistoria europea y para los estudios lingüísticos de los lenguajes indoeuropeos.

El punto que comienza del problema de los orígenes de indoeuropeo no es archaeological excepto lingüístico. Cuando los lingüistas miran los lenguajes de Europa, perciben rápidamente que estos lenguajes son relacionados. Las conexiones pueden ser consideradas en el vocabulario, la gramática y el phonology (reglas para la pronunciación). Para ilustrar el relatedness en el vocabulario, es suficiente comparar las palabras para los números a partir del uno a 10 en varios lenguajes indoeuropeos [ve la ilustración en la página 109]. Tal comparación hace despejado que hay semejanzas significativas entre muchos lenguajes europeos y también Sanskrit, el lenguaje de los textos literarios más antiguos de la India, pero que los lenguajes tales como Chino o Japonés no son miembros de la misma familia.

Más comparaciones detalladas permiten los lingüistas subdividir además los lenguajes de Europa, en las familias. La familia más antigua a ser reconocida, eso de los lenguajes romance - conocidos todo para ser descendido del latín - incluye francés, italiano, español, Portugués y Rumano. Los lenguajes eslavos incluyen Ruso, polímento, Checo, Eslovaco, serbocroata y Búlgaro. Los lenguajes germánicos incluyen alemán, noruego, Danés y Sueco. Se reagrupen estas familias, que ligan las lenguetas lo más de cerca posible relacionadas, lata ellos mismos para formar la familia indoeuropea más grande del lenguaje. Verdaderamente, solamente algunos lenguajes europeos (Húngaro, finlandés y Basque, porque ejemplo) no son miembros de esta familia.

Cómo este arreglo complejo se presentó? Los lenguajes romance sirvieron como el primer modelo para contestar a la pregunta. Aun alguien sin el conocimiento del latín, las semejanzas profundas entre

los lenguajes romance habrían hecho natural sugerir que fueran deribadas de un antepasado común. En la asunción que las características compartidas de estos lenguajes vinieron del progenitor común (mientras que se presentaron más adelante las divergencias, pues divergieron los lenguajes), habría sido posible reconstruir muchas de las características del proto-lenguaje original. En mucho la misma manera llegó a estar despejado que las transferencias condicionales de la familia indoeuropea podrían ser estudiadas y un árbol de familia hipotético ser construidas, alcanzando de nuevo a un antepasado común: Proto-Indo-Europeo.

Esto es el acercamiento del árbol, que fue iniciado en el 1860's antiguo por el philologist alemán agosto Schleicher; abastece que el marco para la manera la mayor parte de los lingüistas históricos piensa aún sobre el desarrollo de las familias del lenguaje. El proceso básico representado por el modelo del árbol es uno de la divergencia: cuando los lenguajes se aislan a partir de un otro, diferencian cada vez más, y los dialectos distinguen gradualmente hasta que se convierten en lenguajes separados.

La divergencia es de ninguna manera la única tendencia posible en la evolución del lenguaje. Solamente una década después de que Schleicher propusiera su hipótesis del árbol, otro lingüista alemán, Johannes Schmidt, introdució un modelo de la "onda" en el cual extensión lingüística de los cambios como las ondas, conduciendo últimamente a la convergencia - es decir, creciendo la semejanza entre los lenguajes que eran inicialmente absolutamente diferentes. En 1939 el lingüista N. S. Trubetskoy de Soviet fue en cuanto a sugerir que las semejanzas entre los lenguajes indoeuropeos pudieran venir sobre esta manera. Hoy, sin embargo, la mayor parte de los lingüistas despiden la sugerencia de Trubetskoy y piensan sobre todo en los términos de los árboles de familia lingüísticos.

End section.

Mr. WALGREN. Okay. All right, well, thank you very much.

When we talk about the Air Force's development of that Russian system, what did that cost? What was the level of effort that went into developing that?

Mr. BOSTAD. Well, there are various figures out; I would say it's in the range of 10 to \$15 million.

Mr. WALGREN. And that was all?

Mr. BOSTAD. That's all.

That cost was basically for our external contracts. We have had an internal cadre that—and its translators, who are sort of part-time, provide feedback to the system. But there has been an internal group that has provided feedback, and that's not factored in.

Mr. WALGREN. How big an internal group, generally?

Mr. BOSTAD. Well, as—at one time I would say in 1978 through 1981 we had eight or nine people working on the Russian system. Now that we have a variety of systems, we've got about five people working on French—French, Russian, German, and some others—and Japanese.

Mr. WALGREN. So it's not an overwhelming—certainly not an overwhelming—

Mr. BOSTAD. No, it's not overwhelming.

Mr. WALGREN. —in any way.

Mr. BOSTAD. No.

I should mention that as a piece of information that the European Commission, which has 15 Systran systems, has invested in the neighborhood of \$20 million in machine translation development.

Mr. WALGREN. I see.

Mrs. Wince-Smith, the conference that just concluded with the report that has been the focus of this discussion this morning, the—did they call for the development of a national plan to develop machine translation technology? Did they say anything specific on that?

Mrs. WINCE-SMITH. That is clearly one of their recommendations that the government has a leadership and catalytic role to play; and I think we've heard that this morning from the representative from the Academy.

And as I noted in my testimony, we feel that this symposium's report lays a foundation for us to begin to do that; and the next step in the process is the JTEC mission to Japan. This is particularly addressing the Japanese case. But I think as we all know this morning that the issue's far broader than just the Japanese to English issue.

Mr. WALGREN. What exactly will be the mission of that JTEC group?

Mrs. WINCE-SMITH. Well, it's my understanding that the mission of the group is to assess the state of the art of Japanese efforts to date vis-a-vis what we're doing and can be doing—and then to come up with some potential recommendations and ideas of how we might work together with the Japanese if we continued to pursue this concept of working together with them, which has been articulated under our U.S.-Japan Task Force.

Mr. WALGREN. Is there a mission statement for them that has been drafted?

Mrs. WINCE-SMITH. Yes, and I think, Dr. Brownstein, you might know some more details about that also. NSF is leading the JTEC effort and Commerce has worked very closely with the Foundation in putting the plan together. And as I mentioned, Dr. Carbonell is the head of it, and he's here with us today.

Mr. WALGREN. Well, maybe we can get him to talk about that.

I guess my reservation is, I would like to see us have a specific plan. And when we talk about international cooperation—although that's called for all the time—it seems to me there's real danger that international cooperation will mean we both just sit there and look at each other; or if there's not a specific contribution that our side is making, we're cooperating for a long time before we realize that we haven't done anything on our own, anyway.

Mrs. WINCE-SMITH. Well, I think we heard, again this morning, that is one of the areas that again we have a strategic advantage in, and I think we'd want to look very carefully about how we would bring that to the table. It's in the whole software systems integration that's going to be critical for developing a machine translation capability.

And if we're going to pursue something with the Japanese we will have to be very careful to see that both sides are bringing to the table shared resources, shared capabilities, and something that balances each other's contribution. And I think that's true for every international project we would want to engage in with the Japanese now.

Mr. WALGREN. What do you see as the advantages and the disadvantages of a cooperative program in terms of what we bring to it? Are there—do we have strengths that we would not want to bring to a cooperative project?

Mrs. WINCE-SMITH. Well, again to speak broader than this particular case, one of the things that we have done some work in the Commerce Department on is that if you look at the number of the international projects, particularly those that are being initiated by the Japanese in the information area, there is across the board a tendency for the Japanese to be approaching us for our advanced work and expertise in the software systems integration area.

That's true in the FSX fighter deal; it's true in the intelligent manufacturing systems area; it's true in the optocomputing, and it would be probably true here in machine translation. I mean, that is a real capability that we have in this country.

And of course at the end of the day we have to decide what we want to provide in terms of what we're going to get that will be in our joint interest, and I think that has to be done very, very carefully.

But at the same time, you know the Japanese, as we know, have invested a lot in this area and I'm sure they have much to bring to the table. It's just a process of working that through the system.

Mr. WALGREN. I see.

You indicate that we'll be looking to this JTEC Study Team for recommendations. We usually don't look to JTEC Study Teams for recommendations.

Mrs. WINCE-SMITH. Well, we look to them for assessments. And I think in this case we want to get an assessment from them of what the Japanese capabilities are; and perhaps using the word "recom-

mendation" is a little broader than is intended here. But we do need to have a first caliber team going in and assessing what the Japanese state of the art is in this area. And that has not been done.

Mr. WALGREN. You really do use the word "recommendations" though, and I want to encourage them to pursue that because I think we need recommendations.

When you say that steps to reach the goal of a viable machine translation await final assessment of recommendations from these groups, and the groups that you're referring to are the JTEC Study Team along with the National Academy of Sciences, which we already have, those are the groups that you refer to?

I guess I'd like to ask, will you be calling for recommendations from those groups as to next steps?

Mrs. WINCE-SMITH. Well, under the U.S.-Japan Task Force on Scientific and Technical Information that Dr. Clark chairs, we have begun a process with the Japanese to talk about potential collaborative efforts in this area.

So the JTEC report, they're very—the study mentioned—the Japanese are very much aware of and know that we're going to be using that as a basis to put together our potential plan. So, you know, however you want to determine it, they will be making recommendations because their assessment will be feeding in to our policy process in the government. And I think that's appropriate, because we can't make a policy recommendation without the underlying technical knowledge and assessment of where the Japanese capability and strengths are in this area.

Mr. WALGREN. Is the government committed to developing this kind of capability?

Mrs. WINCE-SMITH. Well, we've certainly—as I said in my testimony, we've certainly had an accelerated interest in it based on the Academy's Symposium and our own recognition that this is one key path to pursue.

There's no Administration-wide position yet on to what extent the government should be involved in a crosscutting way. And I think this is something that in the not too distant future, once we get all these various pieces together, that is appropriate to bring into the coordinating policy process of the FCCSET, for instance. We have a Technology and Industry Committee that's gotten under way that's doing a critical inventory—doing an inventory of critical technologies and what the government support is at this stage. And machine translation may very well be something we want to look at, because, as you know, right now Commerce has done some support of this through various means, but the two key funding agencies have been the Air Force and the National Science Foundation.

And yet, all of our science and technology agencies have a need for timely state-of-the-art technical information throughout the world to do their own jobs. So I think that this is a potential path that we will be examining.

Mr. WALGREN. What is the level of funding that the Commerce Department is expending in machine translation at this point?

Mrs. WINCE-SMITH. Well, the primary main effort is through our Japanese Technical Literature Program, and we expended approximately \$70,000 for the National Academy Research Council Sympo-

sium. And then we've contributed—we will be contributing \$30,000 to the JTEC mission. And I think you are aware of the budget levels for Japanese Technical Literature Program, a little under \$400,000 at this stage. So that's a, you know, fairly significant amount given the overall task we have to perform under JTL program.

Mr. WALGREN. What other efforts are being made to strengthen the access to the Japanese technical information? What else is happening under Japanese Technical Literature program?

Mrs. WINCE-SMITH. Well, as I mentioned, one of the things we're doing this year that's a new approach and new twist to our traditional support of translating publications, making them available through the newsletters and NTIS, is we're going to be convening a major conference next year with corporate executive types in industry to make them aware of this program, to get beyond the traditional science and technology community.

One of the things that we're concerned about is that no matter how much we do under Japanese Technical Literature Program, if this information only resides in the research entities of corporations, it is not getting into the corporate management system.

And I mentioned some of the programs under way in industry that are very innovative in this context—at MCC, and Motorola where their CEOs are taking the information and utilizing it in a corporate strategy sense. It's not just being utilized down the line by a third tier technician. So this is a marketing effort that we have not really done in Commerce before.

So that's really the new initiative we're pursuing for the next year, is to do a better job of making the possibilities of this aware in the U.S. business community. Because if they don't utilize it, I think we've lost a tremendous opportunity.

Mr. WALGREN. Let me ask you if you would give us some kind of a description of the Japanese Technical Literature program and the—what is being done with the balance of the 400-and-some-thousand-dollars that you have in there.

As you know, there's a widespread feeling that the organization involving that program was something that diminished its effort and made it less effective. And we would like to know exactly what your level in the Administration feels was being accomplished by that program.

Mrs. WINCE-SMITH. All right, I can submit that today. I have that material with me.

Mr. WALGREN. Okay.

Mrs. WINCE-SMITH. The summary of all the various activities that we have supported.

Mr. WALGREN. Okay.

[The information follows:]

Technical Activities: 1987-1990 (Reports and Conferences) of the Japanese Technical Literature Program

Reports:

FY87

Directory of Japanese Technical Resources, 1987 (PB87-205258)
Recent Activities: Ceramic and Semiconductor Sciences in Japan
(PB88-122478)

Report to Congress on the Activities of the Federal Government
to Collect, Abstract, Translate, and Distribute Declassified
Japanese Scientific and Technical Information (PB88-94816)

FY88

Directory of Japanese Technical Resources, 1988 (PB88-211073)
Survey of Supply/Demand Relationships for Japanese Technical
Information in the United States; The Field of Advanced
Ceramics Research and Development (PB88-210943)

Factory Automation in Japan: Key Trends and Innovations,
1988 (PB88-245956)

Seminar on High Tech Material in Japan -Evaluation of
Seminar (PB89-133938)

Current Status of Japanese to English Machine Translation
(PB89-128276)

The Current Status of Science and Technology Grey Literature in Japan
(PB88-227780)

Tapping Japanese Database Resources: An American Strategy
(PB90-163080)

The Japanese Exploratory Research for Advanced Technology
(ERATO) Program (JTECH Report) —

Biotechnology in Japan-1988 (PB89-141147)

FY89

Research and Development in Biotechnology-Related
Industries in Japan (PB89-167936)

Recombinant DNA Research In Japan..A Status Report and
Comparison with the US (PB89-158752)

Basic Research in Superconductor, Ceramic and Semiconductor
Sciences at Selected Japanese Laboratories (jointly with
DOE DOEER-6410)

Assessment of Power Electronics Inverter Technology and Its Impact on the Japanese Air-Conditioning Industry (jointly with DOE)

CAD/CAM for Semiconductor Manufacturing (JTECH Report) (PB89-138259.)

Advanced Sensors in Japan (JTECH Report) (PB-89-158760)

The Japanese Exploratory Research for Advanced Technology (ERATO) (JTECH Report) (PB89-133946)

High Definition Television (JTECH Report with NSF)

Superconductivity (JTECH Report with NSF)

FY90

High Temperature Superconductivity in Japan, Nov. 1989 with JTEC Japanese and American Competition in the Development of Scroll Compressors and its Impact on the American Air Conditioning Industry, PNL-7240 Feb. 1990

Japanese Molecular Biology 1990: An Update, PB-90188707

Directory of Japanese Technical Resources, 1990, PB90-155730

Directory of Japanese Technical Reports, 1989-1990, PB90-163098

Directory of Japanese Databases, 1990, PB90-163080

Intelligent Buildings in Japan

Magnetic Field Facilities Necessary to Characterize Superconductors

Japanese Technology Transfer (JTEC)

Japanese Technical Literature Bulletins Issues 6,7,8 and 9

Proceedings of the Japanese to English Machine Translation Conference, NRC (Dec. 7, 1989)

Translations:

FY88

Research and Development Report on Critical Operations Robots (PB88-204821) (with NTIS)

Study in Biomechnomimetic Materials: Artificial Bone and Muscle (PB88-252325) (with NTIS)

Fuzzy Control Engineering, Toshiba (PB89-128243) (with NTIS)

Science and Technology Agency White Paper, 1987 (with FBIS)

FY89

Intelligent Referencing System for Database of Academic and Technical Literature: Mitsubishi Electric (with NTIS)

IC Card Systems :Toshiba Technical Report (with NTIS) (PB89-232920)

Trends and Tasks of Industrial Technology in Japan (with FBIS)

(JPRS-JST-89-018)
 STA (Science and Technology Agency) 1988 White Paper (With FBIS)
 (JPRS-JST-89-027-I and II)
 Visco-casting Technology, Mechanical Engineering Institute
 (with NTIS)
 CAD/CAM In Japan (NTIS/FBIS)
 MITI Report on Future of Factory Automation (with FBIS)

FY90

Die CAM System Function Specifications, JPRS-JST-89-025
 Future Prospects of FA-From FA to IMS, JPRS-JST-89-026
 Machine Technology Research, PB90-124945
 US Technology Trade Policies and Japanese Responses, (NIRA Report)
 PB90-169699
 Future Prospects and the Results of the Recent Research on
 Production Processing-Technology, Technical Document No. 184
 (Mechanical Engineering Laboratory/MITI/AIST)

Conferences:

FY88

Seminar on High Tech Materials in Japan (Jointly with ONR and
 DOE), April 22, 1988
 Japanese to English Machine Translation Workshop : US
 Government Needs, Sept. 30, 1988

FY89

Superconductive and Semiconductive Ceramic Research in
 Japan (March 29, 1989)

FY90

Dec. 7: Japanese to English Machine Translation (with NRC)
 Dec. 19: Molecular Biology in Japan
 Feb. 14: Advanced Japanese HVAC Technology (ASHRAE-Forum)
 March 21: Advanced Computers in Japan
 April 23: Foreign Manufacturing Technology

Book Reviews:

The following Japanese books and articles were reviewed in the Japanese
 Technical Literature Bulletins as of Sept. 1990

The Future Prospects of FA-From FA to IMS July, 1989 (MITI)
 US Technology Trade Policies and Japanese Responses (NIRA)
 The Japan that can say "No" (Ishihara and Morita)
 Total Prediction for the 90's (Makino)

MITI Industrial Technology White Paper 1988
 STA White Paper, 1989
 Hit Products 1989 (Mitsubishi Research Institute), 1990
 Views on Societal Changes and Technological Revolution (MITI)
 Early Establishment of Reprographic Center in Japan
 Gazing at the US-Japan Relationship (R. Murata-Japanese
 Ambassador to the USA)
 Annual Report of MITI/AIST Laboratory Activities, 1988 and 1989
 Monthly Technical Reports (GIHO) of over thirty Japanese companies
 Monthly Technical Reports of Japanese Government Laboratories
 Monthly Newsletters from MITI/AIST Laboratories

Mr. WALGREN. All right.

Dr. Brownstein, how do you view the mission of this Task Force that NSF is heading up?

Dr. BROWNSTEIN. Are you referring to the JTEC Task Force?

Mr. WALGREN. Yes.

Dr. BROWNSTEIN. Yes, well, we've been through the process, working with Dr. Clark, sitting across the table and discussing what might be pursued under the S&T Agreement with Japan.

Machine translation is one of those things that we intend to pursue. And the JTEC group, like other JTEC groups, would provide us with a better sort of external expert view of what the underlying strengths are, what the state of the art is, where there would be advantages to working together, and disadvantages to working together, and what the barriers would be.

And when we get that information in some systematic form, then we would be in a position sort of collectively to go ahead and develop a firmer plan depending on what they have to say.

I should mention to you that like other such external groups, they're not shy about recommendations. We get recommendations all the time; and we're not shy about receiving them. We'll have to weigh those, you know, along with all the rest of the recommendations that we get. But these groups are given charters to go and try to find out as best they can what the state of the art is, not to pull any punches; to come back and be very unequivocal about what they believe it to be. They make public presentations and they produce written reports, which we distribute.

So there won't be any manipulation or any, you know, contravening of what they have to say. They provide the technical basis for further decision-making.

Mr. WALGREN. Well, it's my understanding that that's right, that they do provide the technical basis and they do the assessments.

What concerns me is that we do not have a plan in any way to go forward. And the Commerce Department is relying on this Task Force to help them develop a strategy on how to use machine translation and to me that's not a plan. And I don't see anybody committing themselves to a plan.

And my question to you is that since they're going to be awaiting the recommendations from the Task Force to develop a strategy on how to use machine-assisted translation, it would seem to me that you would have to tell them exactly what it would take to get from here to there.

And my question is, if they're—as I say, they're relying on you folks to develop—to provide them the basis for a strategy on how to use machine translation—but that implies it exists—and as I understand it, it doesn't exist. And so it would seem to me that in order to satisfy that charge, you're going to have to lay out a way that you can get from here to there. Would you see that within your charter in terms of making recommendations? It's my understanding that goes beyond what the traditional assessment would reach.

Dr. BROWNSTEIN. Well, there's a particular context for their going—their making this visit, and that is to provide us with some better information for putting together what we've agreed to do with the Japanese, which is a joint program of research.

If they go beyond that, that's a different issue. But that's our mission at hand right now. We have agreed to work with the Japanese on a collaborative research effort in machine translation and what we're trying to do is get a good technical grounding for proceeding to develop the specifics of that collaboration.

I hope that you got the distinct impression that there is a lot of good machine translation technology around, and that people, in fact, are making use of it. And the issue of whether to move ahead to get broader use of the systems that exist, to work on development efforts, to produce a technology for market industry to sell, or to do research to produce wholly new, more advanced machine translation systems are a whole collection of issues.

Mr. WALGREN. Let me ask this: As I understand it, you go over there and you assess and you can document the Japanese capability at that point; you're going to assess their capability. And you're going to say that on a—that—towards a goal of 10 units there at 4, and—and we have—we're at 3. But—but still there's 6—there's 6 steps yet to be done.

Now my question is: We can talk about cooperation all we want but somebody has to say what it takes to get from 4 to 10. Now, who's going to say what it takes to get from 4 to 10 other than, well, we'll cooperate and we'll go in that direction.

Dr. BROWNSTEIN. We have inventories of the kinds of scientific and technical problems that you would need to solve to continually press towards what you would think of as 10. We—I don't think at this point we have to go out and get whole new collections of—of undone advanced basic research.

I could tell you—we could tell you what those problems are, of course, the whole range of technologies from optical character recognition to natural language understanding to these pieces of artificial intelligence that you'd want working in that system like that.

Mr. WALGREN. All right.

Dr. BROWNSTEIN. That's not the kind of assessment purely that they're going to do.

Mr. WALGREN. That's my apprehension.

Dr. BROWNSTEIN. The other thing you should recognize is that we have different funding systems in Japan and the United States. The government here has a broad impact on funding to academic organizations and a lot of basic research is done here in academic organizations. The government there focuses a good bit of its attention—

Mr. WALGREN. From what I hear you saying is that you can lay out the problems to get from here—that are necessary to solve to get from here to there.

How can you lay out what's necessary to do to solve the problems? And you can designate the problem but then who will state what it's necessary to do to solve the problems so that we could make some estimate?

I remember John Kennedy saying to the Congress that, look, we're going to put a man on the moon, but if you're not willing to do it, you might as well not even start.

And my point is that somebody has to make an assessment of what's necessary to solve those problems because perhaps we should be saying like John Kennedy said to the Congress, you

might as well not even start unless you're committed to finishing the job and actually doing it, don't even start.

Now, of course, the Congress rose to the bait and they all wanted to start, obviously. But where's that kind of a plan or a road map that actually has distances and assessable quantities of effort? Where is that?

Dr. BROWNSTEIN. There was a time 30 years ago when people would tell you that doing machine translation was a similar problem of putting a man on the moon, that it was an engineering problem—that you had the basic underlying scientific knowledge and it was a matter of applying it and doing engineering systems. I don't think anyone will tell you that anymore.

And when people then are able to pull out parts of that problem and say, these are research issues, they're telling you, we can't give you that information. And that's why it's a research issue and that's why there's basic research needed in three, four, five different areas.

If we had that kind of understanding, we would not have to do the research. We'd be able to go ahead, specify the engineering problem, decide whether we wanted to spend the money, and make a stab at accomplishing the goal.

But that's not the case for getting to that high quality, automated machine translation that has been pulled out as a kind of a broad goal. That is very much the case for continually pressing on and improving the kind of machine-assisted translation systems that exist. And there it's a matter of will and effort to go ahead and build on a considerable amount of work that's been done over the past 30 years—some of which you heard about.

Mr. WALGREN. Mrs. Wince-Smith?

Mrs. WINCE-SMITH. I just wanted to add to what Dr. Brownstein's saying here, that the problem clearly can't be solved in our universities alone; it can't be solved—or should it be solved by government alone. And in the United States system it's not going to be solved by industry alone. It's going to—if we make progress and reach our goal, it's going to be with all three of these players coming together in a major way.

And in terms of developing a plan from the government perspective, what we should do as a government, I think the viable entity and organization to do this is the one that's been set up by Congress—the FCCSET process, where you bring together all the various players in the government and begin to put the pieces together; and we're not at that point yet.

And I think under the new Technology and Industry Committee that we have, that Dr. Bromley set up, would be a perfect vehicle for that. And that's going to be a process that will have to, you know, work its way out. Because, you know, there's a lot—there may be other things that are going on in other agencies that we're not even aware of here today.

Mr. WALGREN. Well, I guess my only hope is, is that out of this mission we may really be able to give some focus on going down that road. I guess I wish someone would lay out how long that road was towards a viable machine translation.

I myself believe is that it would make a critical contribution to unifying the “unifying” is not the word—but to making accessible

a wide range of joint activity between the United States and Japan. And I personally believe that that is critical if so many good things are to happen and so many bad things are not to happen.

So to me it's a very important—there's a very important contribution to be made there.

Some say that, when they look back on this period, they will view the movement of perestroika in Russia as a response to the unification of communication because of the computer, and the velocity of communication. Well, it seems to me that machine translation has a similar kind of contribution in the relationship between the United States and Japan that could open up as critical changes as we have seen in the relationship between the United States and the Soviet Union.

Well, I've gone way beyond my time.

Mr. Chairman, I certainly appreciate your forbearance and your interest in this area. Let me turn the Chair back to you.

Mr. VALENTINE. [presiding] Thank you, sir, I appreciate your helping us out.

Let me—I don't—having missed so much—all of your testimony and questions, I won't undertake to question this panel. But I'd like to make one statement before we call up the last panel.

It occurs to me as a lay person that what we are talking about in this part of the advancement of medical science is the creation of a capability that is always going to have to have some kind of first-class, high IQ, human supervision, is it not? Because who is there to keep the machines from translating *clair de lune* into moonshine?

There are so many different meanings of words in every language. Hasn't there always got to be somebody there to take care of that problem?

[No response.]

Mr. VALENTINE. Okay, thank you. Thank you all very much.

The next and last panel, Panel 3, consists of Dr. Jaime Carbonell, Assistant Professor, Computer Science Department, Carnegie Mellon University of Pittsburgh; Dr. Winfield Scott Bennett, Research Scientist, Linguistics Research Center, University of Texas at Austin; Dr. David E. Johnson, Research Staff Member, IBM, T.J. Watson Research Center, Yorktown Heights, New York, and Mr. Gregory Zaretsky, President of The Corporate Word, Inc., Pittsburgh.

Welcome, gentlemen. Your prepared statements will appear in the record as presented to us, and I request that you summarize.

Dr. Carbonell, I guess you would be first; you're the first person on my list.

Dr. CARBONELL. Is this microphone working?

Mr. VALENTINE. I think it is, if you'd get it a little bit closer to you, sir.

STATEMENTS OF JAIME CARBONELL, DIRECTOR, CENTER FOR MACHINE TRANSLATION, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PENNSYLVANIA; WINFIELD S. BENNETT, LINGUISTICS RESEARCH CENTER, THE UNIVERSITY OF TEXAS AT AUSTIN; DAVID E. JOHNSON, RESEARCH SCIENTIST, IBM CORPORATION, THOMAS J. WATSON RESEARCH LABORATORIES, AND GREGORY ZARETSKY, PRESIDENT, THE CORPORATE WORD, INC., PITTSBURGH, PENNSYLVANIA

Dr. CARBONELL. Thank you.

Just to set the record straight, I'm Director of the Center for Machine Translation at Carnegie Mellon University, where I'm a Professor in Computer Science. And I'd like to make a few brief remarks and then address some of the questions that were asked earlier that I was itching from the back there to be able to answer.

Machine translation came into existence shortly after the end of the Second World War when American researchers who had succeeded at many difficult code-breaking tasks through electrical mechanical devices endeavored to use their skills to translating from one natural language to another—from Russian to English, German to English, and so forth.

These simplistic approaches did not work because there is no one-to-one lexical or even phrasal correspondence between languages. Words have different meanings. When I use the word "bank" I could mean the financial institution, I could mean the side of a river, I could mean an airplane banking to the left, I could mean "I bank on John" to mean that I trust him—and so forth.

As a consequence of errors in the very early translation systems, the philosopher-mathematician, Bar-Hillel, and later the ALPAC Committee concluded that translation, in order to be accurate, required a measure of comprehension, required knowing what the text was about in order to select the right words and the right structure to translate accurately.

And since this was viewed as too difficult back in the 1960s, it was concluded that at that time machine translation was not feasible—it was better to proceed with human translation.

Now this was true back then; we believe it is no longer true now, as many of the efforts in this country and Japan and in Europe are showing.

Perhaps a positive result of the ALPAC report was a return to basic science and computational linguistics in which much of the way of addressing the syntax and the semantics that is the structure and the content of language were addressed and now are being put to use in machine translation systems.

Let me now transition to the present state of the art in machine translation rather than its origins.

The 1990s could herald the long-awaited renaissance in machine translation in the United States. There has been a major hiatus which has not been felt in other countries—in Europe and Japan in particular—because they had no ALPAC report to tell them that machine translation was impossible. They went ahead and did it, with certain degrees of accuracy.

Europe has witnessed a major machine translation effort primarily through the EUROTRA—European Economic Community

effort—which was started in 1978, and sustained a multinational, multimillion dollar effort. Part of EUROTRA was using American technology for translation of documents and part of it—the larger part—was to develop along indigenous systems for translating among the multiple European languages.

There has been significant funding in Europe also from the component countries. That is, the French and the Germans, most notably, have developed their own machine translation efforts. And German machine translation R&D today is 10 times the entire United States R&D effort—governmental R&D effort—machine in translation.

And this is just the German national effort; not counting the French national effort, not counting the European Community efforts, and so forth. Therefore, translation is taken far more seriously offshore, whether it's across the Atlantic or across the Pacific, than it is in this country.

In Japan, the private sector dominates machine translation R&D as well as deployment, with some two dozen companies proudly exhibiting their machine translation laboratories and starting to produce systems for actual use.

The Japanese companies include the major electronic and computer giants: Fujitsu, NEC, Toshiba, Ricoh, even IBM-Japan, Sharp, Hitachi, NTT, and so on.

The Japanese government funds machine translation at various research laboratories, often in collaboration with industry, such as ATR—was mentioned earlier by Dr. Brownstein—the electronic dictionary project, EDR, ICOT of fifth generation, and so on. And even some Japanese universities have done machine translation, most notably Kyoto University where Professor Nagao—known as the father of Japanese machine translation—pioneered the current effort and educated major research figures, who are now scattered throughout Japanese industry. In other words, he provided the seeds to the current MT effort in Japan.

I am afraid that I can count American efforts in machine translation much more rapidly. There are only sort of four or five active research laboratories. The largest one being the Center for Machine Translation at CMU—my university. And this one is supported primarily through private funding, internal funding at CMU, corporations. And much of that funding, over half of it comes from overseas, primarily from Japan, because we are pursuing a different approach to machine translation and they want to, of course, hedge their bets and have access to this method as well as the ones that they themselves are pursuing.

The National Science Foundation funds, I believe about somewhat over 300,000 a year, and at three translation sites at least, a little bit at our site, the University of New Mexico, and at NYU.

And there is a new initiative at DARPA to fund machine translation, which if it reaches its conclusion if it goes through, will add another million dollars a year in government funding to machine translation.

You add all of these efforts together and the total amount is less than any one of the major Japanese corporation spends on machine translation internally at their own laboratories. So the amounts that we're talking about existing at present in terms of government-

tal sponsorship, for machine translation in the United States are reasonable indeed.

In the private sector, IBM is the only large company with a serious machine translation effort. And since we have a representative of IBM, I will not describe what those efforts are in the United States. And some of the small private companies who attempted to do machine translation have fallen upon hard times. The most successful one of them—SYSTRAN—is currently under, I believe it's French, German, and Japanese ownership.

Now I have a few specific recommendations that I would like to make. And these are my own recommendations—they don't necessarily represent those of the other panel members or the other panels.

First, the Federal Government, through the National Science Foundation, or DARPA or other appropriate agencies, should fund the basic research in machine translation. As we've heard before, and I agree wholeheartedly, not all of the research is in place and accomplished, and it is not solely an engineering effort—although there's a large component that is indeed engineering, there is still some basic research to be done.

The Federal Government, however, should not fund targeted machine translation application systems. Those should be funded by the customers. Whether the customers happens to be a governmental agency that needs a task for translation done or private companies who must translate their manuals for overseas sales and marketing, they are the ones that should fund the end use product because they know best what it is that will meet the demand.

However, in order to support this end use, the basic infrastructure must be developed and funded through the government.

And I believe very strongly that all machine translation R&D sites should combine the basic research with the development of end user products. Without completing this lifetime system cycle, there is a high probability that much of the research would go in unproductive directions. I believe very strongly that's the case. That's what we have learned from the Japanese, that development of systems and targeted research working together with the potential users or marketeers of this work usually leads to far more productive use of the funds. And this applies to universities as well as to industrial or governmental sites where machine translation may be developed.

Fourth, there should be something like a half dozen, at least, healthy machine translation R&D sites in the United States. Presently, there aren't that many, for example, the ones focusing on primarily research—only CMU and IBM, I believe, have critical mass at the moment; although there is good work going on elsewhere as well.

Fifth and finally, the overseas machine translation science and technology should be carefully evaluated. The JTEC process that was referred to before, which I'll be leading a team to Japan, is only one small piece of this effort. The European technology must be assessed as well. There is good work there and it must be compared.

And after the assessment itself, recommendations and planning for how to best make use of foreign technology, or collaborate with

foreign sites, should be elaborated and described in enough detail so that we can actually go forth and do this. The JTEC process, as I said, is crucial, but only a piece of the whole pie.

Now, those are my prepared remarks. I had a couple of short answers to previous questions.

One referred to the international collaboration effort, and I would like to add that there are biannual conferences called the Machine Translation Summits, which are held every other year—the first one was in Japan, the second one was in Germany, and the third one will be next year here in Washington, D.C.

These serve as a forum not only to exchange of scientific ideas but also of consumers of machine translation and demonstrations of products or pre-competitive or pre-product technology.

Finally, with respect to spinoffs of machine translation technology, I'd like to enumerate what some of these spinoffs would be. Document gisting—that is, extracting what a document is about in a few short sentences, whether the document happens to be in the same language or in a different language, is a very crucial endeavor; and there are indeed some research projects being funded by government to do precisely that at the moment. Much of the technology is the same basic technology as that that's required for translation.

Document and message routing—get the right document to the right person immediately is another important task, where you've got to analyze the language to know who it is that should read it.

Database query and access. Even if the data happens to be stored in another language, giving it an English query, accessing the right data, and then translating only that which is relevant to provide an answer.

Voice translations through telephone interpretation as we've heard at ATR, and so forth.

These are all related technologies with the same scientific base as machine translation, so we're not talking about machine translation in isolation or in opposition to these, but rather as part of the same family of technologies.

That concludes the remarks I wish to make. At some point, if anybody wishes to see, I have a, I believe it's a 4-minute video tape of language from Japanese into English and voice. And I note the presence of that machine over there.

Mr. VALENTINE. We will certainly take you up on that invitation if time permits.

[The prepared statement of Dr. Carbonell follows:]

Machine Translation Technology: Status and Recommendations

Jaime Carbonell, director
Center for Machine Translation
Carnegie Mellon University
Pittsburgh PA 15213

September 11, 1990

In this brief presentation, I shall endeavor to highlight key aspects of Machine Translation (MT): where its roots lie, the present state of the art, future trends, and some recommendations to improve our national capabilities in this key technology.

The Origins of Machine Translation

Machine translation (MT) came into existence shortly after the end of the Second World War, when Warren Weaver and other American researchers flushed with success at decoding encrypted messages with the aid of electromechanical devices, attempted to apply their techniques to language translation. When translating a Russian sentence, for instance, they considered it to be an encoded version of an English one, and attempted to formulate a mathematical relation between the two, augmented with substitution tables.

Such simplistic approaches were doomed to failure because there is no one-to-one lexical or even phrasal correspondence between languages. Moreover, most words are polysemous; they have multiple meanings and context is required to disambiguate. When I use the word "bank," I could mean the financial institution, or a river bank, or a verb such as "the airplane will bank to the left." In order to translate correctly one must know the correct word sense. In 1960, the philosopher mathematician, Bar-Hillel, wrote that translation requires a measure of comprehension. In those days, computer comprehension of text was well beyond the state of the art, and as echoed in the now-infamous ALPAC report in 1966, MT was deemed infeasible and funding was drastically curtailed in the United States.

The only positive result of ALPAC was a return to the basic science of Computational Linguistics -- essentially the component phenomena of computer text comprehension, such as syntactic and semantic analysis. Shortly thereafter, European and Japanese MT efforts, unhampered by ALPAC, quickly surpassed the early US MT systems, with substantial governmental and private funding. Nonetheless, some private US efforts, such as SYSTRAN and LOGOS, survived the dark ages of American MT.

The State of the Art in Machine Translation

The 1990s could herald the long-awaited Renaissance of MT in the United States. The technology is ready, and the economic necessity for MT is far more pressing than in the early days. Let us review the state of the art.

Europe has witnessed a major MT effort, driven by the undeniable necessities of the multilingual European Economic Community. Particularly noteworthy is the EUROTRA effort, funded by the EEC since 1978, a sustained multi-national multi-million dollar effort. There are also national MT programs in most European countries, most notably France and more recently Germany. German government funding for MT R&D alone, not counting EEC-sponsorship, is presently about ten times the entire United States government R&D funding for MT. Whereas most European MT research takes place at Universities, several companies carry out sustained R&D and commercialization efforts, including Siemens, Phillips, BSO (Holland) and IBM-Europe.

In Japan, the private sector dominates the MT R&D scene, with some two dozen companies proudly exhibiting their Machine Translation laboratories and starting to produce systems for actual use. These Japanese companies include all the electronic and computer giants: Fujitsu, NEC, Toshiba, Ricoh, IBM-Japan, Sharp, Hitachi, NTT, and so on. The Japanese government funds MT research at various laboratories including ATR, EDR, ETL, CICC, and ICOT, backed with a strong technology transfer effort. Some Japanese Universities also perform Japanese MT research, most notably Kyoto University, where Professor Nagao, the father of Japanese MT, started the proverbial ball rolling with his MU system for translation of scientific abstracts. Professor Nagao's students have since occupied scientific leadership positions in many Japanese industrial MT laboratories.

I am afraid that I will cover American MT efforts much more rapidly. There are only four active research labs: The largest being the Center Center for Machine Translation at Carnegie Mellon University, primarily supported by private funding, and over half from overseas. The University of New Mexico and the NYU also have smaller MT efforts. NSF provides some 330K per year to these three sites, and we hope that a DARPA initiative will provide another million a year collectively. The total amount, however, would still be significantly inferior to the budget of a single well-sponsored Japanese MT laboratory.

In the private sector, IBM is the only large company with a serious MT effort. Small MT companies in the US have fallen upon difficult times, with most closing, and others, including the venerable SYSTRAN, falling under Japanese ownership.

RECOMMENDATIONS

In order to improve the state of American Machine Translation, I wish to make a few focused recommendations:

1. The federal government through NSF, DARPA or other appropriate agencies should fund the basic research and development, much more so than the present NSF budget of \$330K per year. As a start in this direction I commend the new DARPA-ISTO program funding interlingual MT research.
2. The federal government should NOT fund targeted MT application systems. These are best funded by the consumers of the technology to be deployed, whether private companies or governmental agencies.
3. All MT R&D sites should combine basic research and technological advancements with development of deployable systems for specific customers. The technology creation, transfer, and deployment cycle should be a closed one. This is a lesson we can learn from the Japanese.
4. There should be a half dozen healthy MT research sites in the United States. Presently there are only two with critical mass (CMU and IBM), and two smaller ones.
5. The overseas MT science and technology should be carefully evaluated, as in the pending JTEC study of Japanese MT technology later this year, and collaboration agreements should be actively fostered with governmental encouragement and support to learn from and contribute to foreign developments in this critical and peaceful technology.

APPENDIX: Present and Future Trends in Machine Translation

The future of machine translation research, development and deployment is the central strategic question which equally interests scientists, engineers, entrepreneurs, corporations and bureaucrats. With a translation market of 20 billion dollars in 1989, growing at a 20% annual rate, MT will remain a vibrant research endeavor and gradually enter the mainstream day-do-day translation market.

Predicting the distant future of machine translation is futile, since the development of any high-technology field is predicated on discoveries yet to be made. A more productive activity is to examine some present trends in MT, especially as they relate to the knowledge-based approach, and extrapolate from them into immediate future. Any extrapolation is, by its very nature, fraught with uncertainty. However, we will try to be conservative, and make conclusions based on the confluence of multiple trends, or those supported by demonstrable pilot research projects in the laboratory. The primary limitation in such conservative projections is most likely to be a horizon effect — not being able to see beyond the current horizon of knowledge, and thereby assuming the risk that unforeseen discoveries, however unpredictable or unlikely, may change the very nature of machine translation.

1. From the Mainframe to the Desktop

Early MT systems necessitated an expensive mainframe; it was the only adequate source of computational power and necessary memory in the 1960s and 1970s. Both commercial systems (e.g., SYSTRAN and the early version of LOGOS) and academic ones (e.g., GETA and Mu) ran on large central computers. Presently, the trend is changing, with the majority of new machine translation systems being developed on single-user machines, such as PCs and individual workstations. Even the KBMT project at CMU with its in-depth semantic processing and consequent necessity for processing power was developed and delivered on IBM PC/RT's.

Not every MT research and development team has yet hopped on the workstation/PC bandwagon, however. The most notable holdbacks, and in retrospect the most obvious ones, are IBM and Fujitsu, the two giants of the computer world. IBM developed and runs all three of its MT systems (SHALT in Japan, their Prolog-based R&D effort, and their pure statistical approach) on large IBM mainframes. Similarly, Fujitsu developed its ATLAS-I and ATLAS-II systems in its own mainframes. The hardware giants, after all, have substantial inhouse mainframe capacity. Their clients both present and future, however, cannot be satisfied with expensive hardware, and we predict that even IBM and Fujitsu will move toward workstations/PCs and distributed computing in machine translation, if they have not already started to march in that direction. This prediction is a fairly safe one. After all, computing in general is moving inexorably towards the inexpensive, powerful, user-friendly networked workstation and PC environment, and MT will ride the wave.

2. From Monolithic Systems to Interactive Symbiosis

The first and second generation MT systems (direct and transfer approaches) were essentially hands-off batch-processing processes. The source text was prepared off-line, then it was fed into the system and hours or days later the target text was given to a human posteditor. Since it was expected that MT output will contain errors, the posteditor was also supplied with the source text, in order for her or him to understand what the target text was supposed to say. The posteditor, typically an experienced translator, would mark up all the corrections off-line, retranslating the more hopelessly confused parts of the text, and these corrections would be later entered, again off-line, producing the final translation. Later improvements, such as the connection of PCs with editing capabilities to the mainframe computer, facilitated data entry and, more importantly, permitted the posteditor to perform corrections directly upon the translated text. But the basic MT process remains to this day essentially a monolithic black box, where human judgment can enter only before input to the central MT process or after output. However, since the quality of most transfer-based translations (exceptions are translations in constrained subworlds such as TAUM-METEO) is not high, human judgment would be extremely welcome *within* the MT process itself.

Recent developments in the transfer-based approach have started to chip away at the proverbial monolith. For instance, a search for unknown words in the source text followed by an interactive lexical entry phase prior to translation is becoming commonplace — cf. the ALEX facility in the LOGOS translation system or the environment in the TOVNA MT system. But, the core problem remains: accurate translation requires comprehension, by human or by machine — or better yet by both if the latter can provide routine comprehension for the bulk of the text, but requires human assistance for the more difficult passages.

This book discussed in some detail how to incorporate semantics into the translation processes in an automated way. Equally important is the question of how to do so in an interactive symbiotic way. The essential idea is for the translation system to make automatically the myriad of low-level semantic judgments (such as matching selectional restrictions to case fillers for prepositional-phrase attachment or simple lexical disambiguation), but to interact with the human user in making the few more difficult judgments. A quick interaction at analysis time may save a much more costly correction down the road (e.g., a translation error with multiple manifestations at post-editing time), especially if the result of the interaction is reused in several places throughout the source text. Moreover, the new knowledge imparted to the system can be accumulated for future use in translating texts in the same subject matter, and thus the system will require progressively less human intervention with accumulating experience.

The notion of interactive disambiguation for machine translation has only recently been analyzed in some depth, and its use for difficult disambiguations in a KBMT system was first introduced systematically in the KBMT-89 project. The automated accumulation of knowledge as a result of online user interaction, however, has not yet been incorporated into a working MT system. We believe, however, that it is only a matter of time, and surely not too much time, before KBMT systems incrementally improve performance through interactive coaching in the true integration of symbiotic user interaction and the knowledge-based approach.

A human posteditor working with erroneous and stylistically deficient target text is essentially performing a janitorial service: cleaning up the mess that the machine made. However, a human interactive disambiguator is performing the role of consultant or tutor, as he or she imparts knowledge that the machine lacks, and receives the satisfaction of observing improved performance over time. This does not totally obviate the need for human postediting as a final verification of the translation (exactly as presently done in better translation agencies with the work of even top-notch human translators), but that role should be much less intensive and intrusive than present MT postediting, where the posteditor is often called upon to retranslate a high percentage of garbled or incorrect target text, and must observe the same errors over and over again. Therefore, we predict that the role of human participation in machine translation will change fundamentally to that of consultant and tutor, and in the long term the student (the MT system) will be largely on its own producing highly accurate and increasingly autonomous translations.

3. From "MAT vs MT" to the "M(A)T" Continuum

In contrast with classical MT, the MAT (Machine Aided Translation) philosophy endeavors to retain human control at every stage of the translation process, providing such help as rapid electronic access to mono- and bi-lingual dictionaries, terminology banks, multilingual word processing, spelling checkers, aligned source and target windows for text editing, immediate access to previously translated passages, and other such conveniences. MAT required the availability of inexpensive desktop word processors and PCs before becoming commercially available, but since it attempted no translation or other computationally-intensive tasks, the earliest PCs and word processors sufficed, giving rise to the first commercial MAT efforts, such as the ALPS system.

Since classical transfer-based MT is a batch process, and MAT is a totally user-driven interactive process, the two have been viewed as opposing alternatives to the partial automation of the translation process. With the advent of KBMT, where human interaction is an integral part of the translation process, the demarcation between MT and MAT blurs into M(A)T. One can envision an entire spectrum of increasingly automated MT workstations, starting from present-day MAT, and gradually automating more and more of the translation process, progressively reducing human intervention until it reaches the point of an occasional consultancy.

In particular, an MAT system can be augmented with increasingly sophisticated analysis and generation functionality such as:

- *Morphological decomposition* to aid in dictionary search by addressing inflectional, derivational, and compounding morphology.
- *Morphological production* to aid in the composition of the target text by conjugating the selected verb, selecting the appropriate declension and gender agreement, compounding words correctly, and so on.
- *Lexical selection advisors* to aid the translator in selecting the most appropriate term in the target language, using context-sensitive corpus statistics such as collocation tables.

- *Syntactic analysis* of the source language to help in the comprehension process if needed, and to provide the first step towards automating the translation.
- *Syntactic verification* of generated target sentences to catch grammatical errors or stylistic deficiencies.
- *Sentence generators* given the representation of a sentence in, for instance, functional structure, render it into target text, possibly querying the user on difficult lexical selections or other problems.
- *Sentential semantic analysis* of the source text to resolve most lexical and prepositional-phrase attachment ambiguities, querying the user to resolve the more difficult residual ambiguity.

As one can see, continuing to grow the list above transitions gradually from a translator-driven process, in which the computer provides essentially clerical background help, to a mixed-initiative task, where the simpler aspects of translation are fully mechanized, to the KBMT method, where the system performs the bulk of the work, and consults the user only on those occasions where its knowledge proves deficient. Moreover, such an MT workstation can be set to different levels of automation as desired by the translator, much as an autopilot can be engaged or disengaged. Even without the autopilot, all the instrumentation still provides much assistance to the human pilot. And, just as in advanced instrumentation for aviation, M(A)T promises not only to improve productivity but quality. If the level of effort required to find appropriate terms is reduced, for instance, it is much more likely that the translator will exert that effort rather than be satisfied with a "somewhat acceptable" alternative in the interest of expediency. If the system requires correct grammar on output (by generating it automatically, or by checking the translators' output), a typical source of error is removed. M(A)T, therefore, represents a gamut of possibilities in degrees of automation.

The scenario briefly sketched above is not yet a technological reality; it is our projection that it will come to pass in the not-too-distant future. Laboratory work in this direction is already underway, and the model of variable-automation in the translation process is gradually being accepted as a viable alternative to either of the extreme positions — full MT or rudimentary MAT.

4. From Knowledge-Free to Knowledge-Based Machine Translation

Machine Translation in its early heyday, prior to the harsh realities of the ALPAC report and Bar Hillel's assessment that translation requires at least a modicum of comprehension was viewed primarily as a transduction task. Sentences in the source language were merely character strings to be decoded into the corresponding sentences in the target language. And, this process was to be done via simple transformation functions (augmented with lexical correspondence tables), paralleling the encryption/decryption technology for military messages developed during World War II. As discussed earlier in this book, this *direct MT* approach failed to take any principled account of linguistic structure or semantic content, and was by and large relegated to dustheap of

unsuccessful approaches that, as mathematical elegance dictated, "should have worked" but did not.

However, the spirit of doing as little linguistic and semantic analysis as possible survived well beyond ALPAC in the MT community. After all, who can blame researchers faced with such a seemingly insurmountable problem: full analysis of arbitrary text is arbitrarily difficult. Therefore, the name of the game became one of performing the minimal amount of analysis to permit transfer to the target language, and therein the modern transfer-based MT paradigm was born. Over time, however, the analysis process became more and more comprehensive in order to improve the low accuracy of early MT systems. And, some improvement was indeed noted. Most modern transfer approaches recognize the need for in-depth morphological and syntactic analysis – but not yet the equally important need for in-depth semantic analysis. This final barrier has proven more formidable for scientific and perhaps even psychological reasons. With respect to the latter, at the first Conference on Theoretical and Methodological Issues in Machine Translation, Victor Raskin remarked on the hesitancy of MT researchers to address semantics head on, and coined the word *semantophobia* – "fear of meaning." Full semantic analysis is indeed an extremely knowledge-intensive and cycle-intensive process, but we prefer to view it as a great challenge, rather than as a reason to run in the opposite direction.

The first attempts to perform systematic semantic analysis were confined to the the interlingua MT paradigm. After all, in order to represent the content of the text in an intermediate representation, the interlingua, semantic analysis was an indisputable requirement. Tradeoffs in depth of analysis vs. breadth of coverage surfaced, and the first interlingua systems were confined to restricted domains, that nonetheless grew larger over time. Some efforts such as ATLAS-II at Fujitsu and ??? at NEC attempted an interlingua approach with a shallower but broader sentential semantics coverage.

Although the proponents of the interlingua approach, present authors included, have argued the need for in-depth semantic analysis, that need can be fully recognized and decoupled from the transfer-vs-interlingua debate. That is, transfer systems require semantic analysis for accurate translation just as much as interlingua systems, even if the former have no intermediate "pivot" representation. With considerable reluctance, the transfer-based MT projects are indeed starting to add an initial measure of semantic analysis, but still fall far short of the depth of analysis we believe required for accurate MT. One important example is the argument for feature-based semantic processing in the transfer-based METAL translation system (Bennett, 1990). Manifold increases in computing power and memory capacity have reduced significantly the run-time efficiency penalty for in-depth analysis, whereas the accuracy penalty for shallow analysis remains forever with us.

We believe that the trend towards increasing depth and sophistication of source-language analysis will continue at an accelerated pace in the coming decade. Driven by quality objectives, many of the better transfer-based MT systems will perform increasingly deep semantic analysis. And, the interlingua systems will continue to view in-depth semantic analysis as a crucial component in full multilingual MT. Although we do not predict convergence on the transfer-vs-interlingua dimension, we do predict steady progress towards deeper and more thorough analysis – essentially an acceleration of the historic trend from knowledge-free to knowledge-based machine translation.

5. From Barely Intelligible Output to Fluent Translations

Two typical criteria for evaluating machine translation are accuracy and intelligibility. The former one measures the semantic fidelity of the translation, the degree to which the target text preserves the meaning of the source text. The latter one measures the degree to which the target text is well-formed and comprehensible, without postediting or clarificatory access to the source text. In simple terms, these criteria may be paraphrased as follows: "A translation is accurate if it says the right thing, and it is intelligible if it says it the right way."

By and large, this book addresses the contribution of KBMT to improving the accuracy of the machine translation process. And, as discussed above, accuracy is largely a function of the depth of analysis, rather than a function of the translation paradigm, although interlingua approaches typically involve a deeper semantic analysis than pure transfer ones. However, improving the intelligibility of target text is an important criterion, and one facilitated by the KBMT approach. If the analysis of the source language provides information beyond the propositional content of each sentence (such as topicalization, speech act, politeness, and so on), such information can be used to generate higher quality output text. Both intelligibility and appropriateness of the target may be improved with proper use of such information.

In addition to starting out from a richer knowledge structure there is a recent emphasis on developing powerful target language generators, in contrast with the recent past, when most of the MT research focused on the transfer and analysis phases. The introduction of principled lexical selection rules and text planning beyond the sentential level are providing the necessary ingredients for significant improvements in the intelligibility of automatically translated text. We project that these improvements will migrate from the laboratory to production-use MT systems, and that the continued emphasis on natural language generation will yield additional intelligibility improvements.

6. From Mono-Media to Multi-Media Machine Translation

Although the primary commercial requirement for MT is to translate significant volumes of printed text, there is an acute need for simultaneous translation of spoken language. Speech recognition technology, coupled with language analysis is starting to emerge from the laboratory. Such technology permits translation of spoken language in domains limited by vocabulary size, semantic complexity, and relatively clean acoustic input. The additional constraints imposed by speech recognition are twofold. First, any recognition system will make acoustic recognition errors; even the well-tuned human ear does not recognize words with 100% accuracy. Second, spoken language is necessarily unedited. Sentence fragments, restarts, corrections, mispronunciations and grammatical errors abound. Lexical, grammatical and semantic constraints are called upon to help select among different candidate words hypothesized by the acoustic recognizer from the input signal. The tighter these constraints the more acoustic recognition ambiguity may be tolerated. And, for spontaneous speech, KBMT is a requirement as noted even by the proponents of transfer-based minimal-semantic systems. When faced with weaker grammatical constraints, a strong semantic model must be present to take on the task of resolving recognition ambiguities.

At present there are no viable commercially-available speech-to-speech MT systems, but several laboratories around the world are working very hard to produce them (ATR, NEC and Matsushita in Japan, and CMU in the USA, for instance). But, speech recognition technology is rapidly improving, partially driven by faster hardware to feed its voracious appetite for computing cycles. Speech synthesis, on the other hand, is simpler than recognition if one does not place demands for subtle intonation changes and other prosodic phenomena. We expect that for well-defined tasks, such as telephone assistance, airline reservations and office automation functions, speech recognition integrated with MT will be developed in the near future. General-purpose speech-to-speech MT, however, must await continued scientific progress well into the twenty-first century.

In addition to pure speech-to-speech MT, one can envision other translation modalities, such as speech-to-text MT, text-to-speech MT, and mixed-mode MT, where both source and target can be a combination of textual and spoken language. In particular, as hypermedia moves from the laboratory to the office environment, documents will combine text, diagrams, images, speech, and executable programs. Only the language-oriented parts of these documents will require translation: text, speech, labels on diagrams, output messages from the programs, etc. However, these translation tasks must be unified, as context from one can constrain the analysis of another. At present, there is no large-scale effort in multi-media MT, as multi-media documents are just starting to emerge and become accepted in the most pioneering of environments. But, we expect that multi-media MT will not trail the diffusion of multi-media documents by very much time.

7. From Bilingual to Multilingual Systems

Early MT efforts considered only translation for a directed *language pair*. Modern transfer approaches have not changed this philosophy; MT companies talk about adding a new (unidirectional) language pair as a momentous event. If translation is required only among two or three languages, a language-pair philosophy make sense, but not so for translation among many different languages. There are $N(N-1)$ (almost N^2), language pairs among N languages. Consider the European Economic Community, where there are 10 official languages: Danish, Dutch, English, French, German, Greek, Italian, Portuguese and Spanish. To translate official documents among all languages, as legally mandated by EEC regulations, it would require 90 transfer grammars! If the EEC were to expand to include Austria, Scandinavian countries, Eastern Europe and Turkey, the number of languages may well double, and the number of transfer grammars would quadruple. Each such grammar will have to be painstakingly developed with person-decades of effort. Of course, Europe is not the entire world. If one adds only the major Asian languages, the number of languages would double or triple again, requiring *thousands* transfer grammars.

In order to minimize the complexity of the language-pair dependent transfer process, some MT systems, including the old and venerable SYSTRAN, attempt to maximize the sharable monolingual analysis process. In other words, these transfer systems perform a degree of analysis of the source language independent of the target language, before applying the source- and target-dependent transfer. Such measures ameliorate the problem, not by reducing the number of transfer grammars, but by reducing somewhat the size and complexity of each transfer grammar. The interlingua approach takes this process to the extreme, analyzing the source language into a semantic encoding,

the interlingua, independent of both source and target languages. The transfer component is thereby reduced to the null operation. The number of grammars needed for the interlingua approach is *linearly* dependent on the number of languages, requiring a single analysis and a single generation grammar for each language. Moreover, the analysis and generation grammars of a given language may share most of their rules.

What exactly is our prediction for multilingual translation? Will interlingua become the dominant paradigm in a few short years? Let us reexamine the basic trends in MT:

- *Increased demand:* The overall volume of textual material in need of translation is increasing inexorably, partially driven by commercial needs (e.g., by globalization of markets), partially by international organizations (e.g., the EEC), and partially driven by the information revolution (e.g., on-line network access to volatile and archival information).
- *Increased linguistic diversity:* Translation is needed among a growing pool of languages.
- *Increased low-cost computational resources:* The techno-economical barrier to widespread deployment of MT systems is coming down with amazing rapidity.
- *Increased electronic integration:* Multiple tasks are being integrated such as: authoring of manuals, document production, multilingual translation, and maintenance and revision. Or, in information gathering tasks such as intelligence analysis: document entry (OCR, conversion from other electronic forms, etc.), filtering for relevance, automated routing, content extraction for database fills, and translation from diverse source languages.

These trends make a convincing argument for the inexorable expansion of the role of machine translation, but not necessarily for one specific paradigm or method over another. They do, however, hint at the universal bottleneck recognized by all MT researchers: the development time and effort required to add new languages, extend existing ones, improve accuracy to reduce or eliminate postediting needs, and so forth. The methods most efficient in terms of minimal development effort and maximal quality of translation will triumph in the survival of the fittest. Run-time machine efficiency is less a concern than in was the mainframe behemoths of yesteryear; the cost equation is rapidly changing with 100-MIPS computing power appearing on the desk top at prices not much more than top-of-the-range PCs.

We have already projected the need for increasing depth of analysis on the basis of improving translation accuracy, with the trend that MT will move more and more towards the knowledge-based approach. Now, we reinforce this projection from the software engineering standpoint: maximizing the reusable components of the translation system (those that do not depend on the cross-product of source and target languages). Hence, the analysis and generation phases will take an increasing proportion of the overall MT effort and the transfer component will be minimized. Interlingua approaches will continue to flourish, and take on an ever-increasing role in multilingual MT, but engineering improvements and minimization of transfer components will keep those transfer-based systems that take analysis seriously in business for a long-time to come.

8. From "MT vs NLP" to Multifunctional Modules

Natural Language Processing (a.k.a. Computational Linguistics) was born from the ashes of the first machine translation wave in the 1950s and early 1960s as a back-to-basics movement. The realization that translation requires a much higher degree of linguistic sophistication (grammatical, morphological, lexical, semantic) than initially foreseen led to the investigation of computational approaches to processing language: parsing, generation, representation, and so on. Natural Language Processing (NLP) drew apart from machine translation over time, focusing on issues such as computational theories of parsing, analysis of discourse phenomena, and interactive monolingual applications such as natural-language database query. In part, the stigma associated with MT by the ALPAC report caused many American computational linguists to stay away from MT. Most early NLP research was done only in English, although more recently work in Japanese, French and German abounds. Moreover, the segregation of MT and other NLP research was aggravated by the builders of MT systems, many of whom viewed computational linguists as raving theorists. [Surprisingly, theoretical linguists viewed computational linguists as impure, applied computer hackers.]

A subtle reconvergence of NLP with MT is already underway, in part fueled by the rise of NLP in European and Asian countries, where there was no damaging ALPAC report, and where MT is a more recent and flourishing activity. Even in the United States the long shadow that ALPAC cast on MT has finally been lifted. In addition to the sociological phenomena, there are sound scientific reasons for the reintegration of MT and NLP. MT is the paradigm task for all of computational linguistics. No application is more demanding than accurate, multilingual MT; no computational model of language can be labeled as comprehensive if it does not address the all-important needs of MT. Accurate, multi-lingual MT is becoming the ultimate *stress test* for NLP.

The advent of modern approaches to MT, KBMT centrally included, are based on advances in computational linguistics, such as unification grammars, frame-based semantics, and efficient parsing methodologies. Computational linguistics is devoted to deep text analysis, and the trend in MT is precisely to deepen the level of analysis in order to improve translation accuracy. The work on MT-specific transfer components is declining as a fraction of the total MT effort, even in transfer systems, as discussed above. Hence, the forces of reintegration are starting to be recognized by MT and NLP researchers to their clear and mutual benefit.

In addition to the overall reintegration, we predict another trend, not yet in ample evidence: multilingual, multi-functional natural-language processing modules. As MT nudges NLP towards a multilingual orientation, it becomes evident that NLP components such as parsers and generators can indeed be built to work in many different languages by decoupling the language specific components (grammar and lexicon) into declarative knowledge bases. Indeed, paradigms such as unification-based parsing are extolled for their multilingual virtues (cf. Kaplan et al., 1989). Moreover, as the analysis phase in MT systems deepens, it matches the parsing task required by other NLP applications such as interactive natural language interfaces to data-base managers, expert systems and other software packages. Development time and effort can be offset better if parsers, generators, morphological analyzers and other NLP modules are developed with multi-functional as well as multilingual objectives.

The underlying processes of language comprehension and generation are basically independent of application domain; it is a second-order effect whether one is composing text, translating it, or reading it to extract basic content. Humans do not learn English specifically for translation and then start all over again with English for reading newspapers. Rather, the basic language skills are acquired once, and then specialized (often with additional training) for important tasks such as translation. We project that the same will be true by the turn of the century for NLP systems. First, the basic processing modules will be developed and reused for multiple functions. Then, the linguistic knowledge sources for each language will be codified and used in multiple applications. Finally, the systems will be reconfigured, optimized and extended when necessary to satisfy the demands of the specific application domains. Reusable multi-functional and multi-language modules will be the accepted mode of operation, and it will be a source of puzzlement why it has not always been so.

Mr. VALENTINE. Dr. Bennett?

Dr. BENNETT. Thank you.

I, rather than reading my paper, I want to just make a few remarks on it. One essentially presented is a case study for what's called the METAL machine translation system. And this is not meant to be some kind of boosterism, but rather to highlight one program that's been going on in the United States for about 30 years.

Work at the University of Texas started in 1959 with support and encouragement from the Georgetown group. We continued on through the ALPAC period, and in fact were one of the few survivors of the ALPAC report. If you look at the paper you can see sort of a summary of the kinds of support we got. As has been pointed out earlier, the report—the funding is entirely Department of Defense and NSF.

We continued on basically to 1974, at which point our funding basically was from the University, except for a brief—about 18-month—hiatus where there really was no work at all—work continued on.

In 1978, the Rome Air Development Center (RADC) funded a feasibility study for creating an operational system.

I should mention, along the way the machine translation effort at the University of Texas got the name METAL, which apparently stands for meta linguistics. I say apparently because there's some debate even among us.

In 1979, Siemens, a large German company, became interested in machine translation primarily for the translation of a large number of documents. They began funding of our work and with their interest our work turned from theoretical research to more development of a production system. They've been funding us since 1979. I joined the project in 1979 and have been instrumental in creating the METAL system.

Basically, METAL is two things. On one hand it's a frame for machine translation. We think that it's theoretically well motivated, therefore, we have continued as a research project.

On the other hand, it's a production system, now comprising four marketed language pairs and we're about to go to a fifth.

The emphasis, however, has been primarily to market the system in Europe, partially because it's a European company that now provides all the funding. And the other being an assumption that there really is a market in the United States.

This ties into Dr. Vasconcellos' earlier comments about work in the United States and perceptions of markets, et cetera.

Our—as I said, we have an international effort now. The work is not only at the University of Texas but is also in Munich, Germany; Leuven, Belgium; Barcelona, Spain, and most recently, Kolding, Denmark.

The difficulty we face is—and here I completely agree with Dr. Carbonell—our efforts are more now towards production systems because we're market-driven. The work on METAL primarily focuses on where our answering market needs, and primarily market needs in Europe.

So the language pairs we we end up working on are the language pairs that we can sell. It's really quite simple.

As far as R&D money beyond what we get from Siemens, we don't have any. So what we have become is a U.S. product, a U.S. technology, that now is essentially owned by a German corporation and marketed by that corporation.

If it is marketed in the United States it will be by Siemens Corporation.

I should add that not only am I a research scientist at the University of Texas but I'm also a senior AI analyst for Siemens-U.S.

One of the points I'd like to make about the system, in answer to some of the earlier comments about machine translation, machine-assisted translation, we are, we believe, a cost-effective system. There are about 25 sites where we have installed a range of translation efforts are from simple—are from translation bureaus to computer companies, to a detergent company, to several automotive companies—all of this in Europe.

Turning from this sort of brief discussion about METAL—as I said, as an indication of a research project that's been going on now for 30 years, essentially, I'd like to address the idea of the Federal role.

I believe that, as other people have stated before, the—this committee, the Federal Government, should look into encouraging a collaborative effort in much the way the Japanese do it and much the way it has been done in Europe, with the government, with industry, and with universities, for furthering various kinds of machine translation.

I'd like to second Dr. Carbonell's suggestion about various MT sites. As he says, essentially there are two now—his own at Carnegie Mellon and IBM. I would say we're a third MT site except that we don't have any research money for pure research, for all our effort, as I said, is now focused on product development.

I also think that any kind of collaborative effort should encourage looking at wholly new approaches to machine translation. As I point out in my paper, citing Dr. Somers from the University of Manchester Institute of Science and Technology, the tendency is to work on the same old theme, and I think it's probably time, as the Japanese and some European groups have been doing, to start looking at other possibilities for machine translation.

Finally I'd like to say that I think that trying to look at MT as a single approach or a single phenomena gets us into trouble and was perhaps the kind of trouble we got into before the ALPAC.

The reality is there is a market out there and the reality is that market is quite diversified. Various approaches to machine translation seem to be appropriate.

Mr. Bostad told us earlier about their successful use of SYSTRAN, which is an old system and for which the conventional wisdom has it that it's not a very good system. But Mr. Bostad and their experience proves that that is not true. Clearly, there are other markets, more personal uses of systems.

Our system, for example—METAL—is geared only for high volume, large corporation use. It certainly wouldn't be something you would ever put on your IBM PC or your McIntosh, or something like that. But I think that any kind of collaborative effort involving the government industry and universities should also look

at possibilities of different kinds of machine translation and machine-assisted translation.

Thank you.

[The prepared statement of Dr. Bennett follows:]

The METAL Machine Translation System

Winfield S. Bennett, Ph.D.
Linguistics Research Center
The University of Texas at Austin

1. Introduction

The University of Texas at Austin has conducted research in machine translation (MT) since 1959. With the exception of a brief hiatus in the mid 1970's research in this field has continued without interruption for over thirty years. The research at the University of Texas at Austin was even able to survive the virtual elimination of MT work in the U.S. following the publication of the ALPAC report.

The Linguistics Research Center (LRC) at the University of Texas at Austin was founded in 1961 as a center for MT research efforts. Dr. Winfred P. Lehmann is the founder of the LRC and has been the only director throughout its long history. In the course of more than thirty years of the MT studies the Linguistics Research Center has obtained quite a bit of expertise in the field.

2. History

Funding for the MT research at the LRC has come from a variety of sources. Although work over the years has generally focussed on the German-English language pair, the LRC's expertise has also been applied to other languages. Some of the work was done in conjunction with other institutions.

The following table presents a summary of the projects at the Linguistics Research Center between 1959 and 1974:

DATES	LANGUAGE PAIR	SPONSOR
1959-62	German-English	U.S. Army Electronics Command
1961-62	German-English	National Science Foundation
1962-63	German-English	U.S. Army Electronics Command National Science Foundation
1963-64	German-English	U.S. Army Electronics Command National Science Foundation
1964-65	German-English	U.S. Army Electronics Command National Science Foundation Wright-Patterson AFB
1965-66	German-English	U.S. Army Electronics Command National Science Foundation
1966-67	Russian-English	Rome Air Development Center
	German-English	U.S. Army Electronics Command National Science Foundation Rome Air Development Center A.F. Office of Scientific Research
1967-68	Russian-English	Rome Air Development Center†
	Chinese-English	Rome Air Development Center
	Russian-English	Rome Air Development Center†
	Chinese-English	Rome Air Development Center
	German-English	Rome Air Development Center
1968-69	German-English	Rome Air Development Center
1969-70	German-English	A.F. Office of Scientific Research
1970-71	Feasibility Study	Rome Air Development Center
	German-English	Rome Air Development Center
	German-English	National Endowment for the Humanities
	German-English	National Science Foundation
1971-72	German-English	Rome Air Development Center
	German-English	National Endowment for Humanities
	German-English	National Science Foundation
1972-73	German-English	Rome Air Development Center
	German-English	National Science Foundation
1973-74	German-English	Rome Air Development Center
	German-English	National Science Foundation

† Two projects: one separate and one jointly with Georgetown University.

The amount of funding during this period varied somewhat from year to year with an average of approximately \$324,800 per year for the years 1962-1974. Funds from the U.S. Army Electronics Command, Rome Air Development Center, and the National Science Foundation tended to be more than \$100,000, while the others were less than \$50,000 per grant. The name METAL, which has been used for all the MT systems developed at the LRC in the past one-and-a-half decades, was first used in the early 1970's.

During the period between 1974 and 1978 work at the LRC was supported by small grants from the University of Texas at Austin itself. There was an eighteen-month interruption during the 1970's. In 1978 the Linguistics Research Center was given a contract by Rome Air Development Center (RADC) to study the feasibility of an operational German-English machine translation system. The RADC study was completed after thirty months in October 1980; the results of this study were published in *The METAL System* (RADC 1981).

In December 1978 Siemens AG, a large German electronics firm, began supporting machine translation research at the Linguistics Research Center. The immediate goal was the creation of an automatic high quality MT system. Siemens' interest in machine translation stemmed from its need to translate approximately 100,000 pages of telephone switching manuals from German into English. Siemens also sponsored work on German-English MT for LOGOS in the late 1970's and early 1980's.

Siemens' initial investment in METAL research was fairly modest: less than \$182,380 for the first thirty-three months (1 December 1978 - 31 August 1981). The financial support increased significantly beginning 1

September 1981; Siemens' support has continued to the present. During this period Siemens has contributed an average of approximately \$194,700 to the University of Texas at Austin per fiscal year. This figure for funding is misleading, however, since Siemens has also provided personnel both on-site and off-site in support of the research, as well as money for equipment, equipment maintenance, travel, and miscellaneous other expenses throughout its entire association with the LRC.

Work on a production version of METAL German-English was initially confined to the LRC, with support from Siemens personnel at a company site in Boca Raton, Florida. From October 1980 on Siemens has been essentially the sole support for the research and development of the METAL system at the LRC.

During the period of 1979-1985 METAL developed from a conceptual model to a framework on which to build translation systems for specific language pairs. German-English was the particular language pair used for research throughout this period, but the task of creating the framework was at least as important as producing translation systems for particular a language pair. Refinement of the framework continues to be a significant part of METAL research.

In fall 1985 the framework was exported to Belgium for the development of Dutch-French and French-Dutch machine translation systems. This work is centered at the University of Leuven, with sponsorship from the Belgian government and Siemens SA, Belgium. Research work on the framework and some support work for German-English began in Munich, Federal Republic of Germany, in early 1986. An effort to develop a German-Spanish pair began at Siemens SA in Cornellá (Barcelona), Spain, in the fall of 1986. The transfer of the German-English system from

Austin to Munich took place in fall 1987; since then, the Austin site has focussed on English-German. Work on German-Danish translation began in 1989 at the Handelshøjskole Syd in Kolding, Denmark. All sites continue to work on improvements to the overall framework as well as on the details for particular language pairs.

Marketing of METAL German-English began in 1988; marketing of English-German and German-Spanish began in 1990. The Belgian Dutch-French and French-Dutch systems are scheduled for marketing either late 1990 or early 1991. To date there are more than twenty installations of METAL systems in Europe. Plans for marketing in the U.S. are underway, but details are somewhat uncertain at the present.

3. The Present System

METAL is an automatic high quality translation system. The system is designed is to be a high volume non-interactive system to which the user sends his or her source language texts and from which he or she gets the target language output of the translation process. Use of post-editing is expected, if for no other purpose than to validate the results of the translation process.

METAL is a highly modular system in which the analysis modules and the synthesis modules are independent of one another. This allows the rapid development of new language pairs, using existing analysis and synthesis modules. For example, the German analysis module from German-English is also used for German-Spanish and German-Danish. The English synthesis module from German-English is being used for a French-English pair currently under development in Belgium.

METAL uses an augmented phrase structure grammar for analysis coupled with a framing mechanism which uses the valency of a given verb to aid in analysis and to determine the functions of the elements of clauses. The analysis of a given sentence is deeper than in most commercial systems, although a 'minimalist approach' is used to provide enough analysis for purposes of translation without delving into the deepest possible structure. Grammar rules are ordered on the basis of their significance to the system.

METAL monolingual dictionaries are used for both analysis and synthesis, i.e., a given dictionary entry serves both for source or target language purposes. The project has always sought to make the dictionaries optimally codable by naive system users. System users are expected to build and maintain their own user-specific dictionaries.

Although METAL uses a traditional transfer (i.e., bilingual) dictionary, the system has decided interlingual tendencies. Rather than being characterized as a transfer system, as it was initially conceived, METAL is now something of a hybrid transfer-interlingual system. Currently work is underway at all the sites to produce a standardized METAL Interface Representation (MIR) for all analyses. Once established the MIR should facilitate the connection between a given analysis and any syntheses.

In semantics METAL takes a 'minimalist approach' as well. Semantics are represented in terms of primitives coded as values for the appropriate feature of a given lexical category. The use of semantics in METAL is viewed as an enhancement to the overall system, particularly to the syntactic analysis, rather than as an end in itself.

To facilitate the acquisition of lexical items (dictionary entries) the LRC has developed a major lexicon tool, the Intercoder. The Intercoder

allows users to enter dictionary items with a minimum of effort. To facilitate the development of the grammar and grammar-related points, the METAL Shop provides a suite of tools for linguists to develop the grammar and related items.

4. Future Prospects

The increased interest in machine translation both in the U.S. and abroad coupled with the maturity of the field has engendered considerable activity during the past five years in improving existing approaches and in considering new models for the field.

Clearly work must continue along the familiar lines of the past years. Even well-defined, established systems must be continuously upgraded, sometimes drastically. New technology as well as new linguistic insights permit and even demand changes to systems. AI techniques likewise open up possibilities for system improvements. For example, METAL's framework is based on a standard transfer approach, as are most commercial systems; however, as indicated above, work is now well underway to modify this system to be more of a hybrid transfer-interlingual system.

A number of new approaches to machine translation have been proposed in the past few years. Dr. Harold L. Somers, a professor at the University of Manchester Institute for Science and Technology (UMIST), recently listed a number of new approaches which he characterizes as 'some research projects ... which ... have in common that they reject, at least partially, the orthodoxy of the "second generation and derivative" design, or in some other way incorporate some ideas which I think significantly broaden the scope of MT research.' (Somers 1990, 5). Among the new ap-

proaches Somers describes are: approaches in which the focus is on a clearly and expertly defined domain for translation; systems which can revise their own grammars on the basis of input from post-editing; systems for monolingual users; dialogue machine translation; and statistical approaches.

5. Federal Role in MT Research

The reality is that machine translation research is underfunded in the U.S. This is a general statement, without reference to the amount of funding available in Japan, for example. This current situation is, I think, the result of a combination of at least two views: (1) a general assumption that such research should be funded entirely by the private sector and (2) a genuine concern about machine translation research stemming from the ALPAC report. I will focus on the first view, since I believe that the concerns raised by the ALPAC report are diminishing with time and increased expertise in the field.

The reliance on the private sector for financial support for MT research ignores the reality of such research. Machine translation systems are not short-term projects. METAL, as the history related above indicates, represents eleven years of work (1979-90), which is itself based on twenty years of work (1959-79). I believe that Siemens' commitment to this work is commendable and unusual for a western private company. Where it exists at all, most private sponsorship in North America is for only short-term theoretical projects which provide useful theoretical models, but never really result in large-scale systems. Japanese commercial support is clearly different, as the report of the symposium sponsored by the

Office of Japan Affairs clearly indicates (Office of Japan Affairs 1990, 32).

None of this is meant as a criticism of the U.S. private sector. As the Office of Japan Affairs report states: 'For the large companies with big R&D budgets ... the high risks and uncertain returns of expenditures on machine translation research present significant disincentives.' (Office of Japan Affairs 1990, 34)

Current funding for MT research has led to another situation. Just as Siemens has provided support for METAL, Japanese and other non-U.S. companies are supporting machine translation research in the U.S., with the result that the profits from technology developed in the U.S. is made by non-U.S. companies.

This leaves the public sector, specifically the U.S. government, as a source for funding for machine translation research. There is apparently more U.S. government interest in machine translation now than ever before in the post-ALPAC era, if the recent DARPA announcement of limited funding for the development of interlingual prototypes is any indication. Such renewed interest is heartening; however, the words *limited funding* must be emphasized, since my understanding is that DARPA will fund at most two such projects for up to five years. This is a start, but only a start.

In my view support for MT research should not be only laid on the shoulders of the U.S. government. I would echo the words of the report, cited above (Office of Japan Affairs 1990, 34), when it recommends a joint enterprise combining the private and public sector, but the initiative should be for more language pairs than just Japanese-English. Besides work on large-scale systems for particular languages, I would further

urge support for the sort of research mentioned in the previous section, i.e., refinement of established existing approaches and exploration of innovative ideas.

Clearly the interest of this committee is a first step, but only a first step. The U.S. government must take an active part in assisting machine translation research in the United States, if we wish to participate in the technology.

REFERENCES

- Office of Japan Affairs. 1990. *Japanese to English Machine Translation. Report of a Symposium*. Washington, DC: National Academy Press.
- RADC. 1981. *The METAL System*. RADC-TR-80-374. Griffiss AFB (New York): Rome Air Development Center.
- Somers, Harold L. 1990. Current Research in Machine Translation. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*. 1-12. Austin (Texas): University of Texas at Austin.

Mr. VALENTINE. Thank you.

Dr. Johnson?

Dr. JOHNSON. Good morning, Mr. Chairman and members of the subcommittee. My name is David Johnson. I am research scientist at IBM's T.J. Watson Research Laboratories in Yorktown Heights, New York, where I am involved in research on machine translation.

This morning I will provide a summary of IBM's work in this important area and, of course, will be glad to answer any questions you might have.

Our interest in machine translation is in part a self-serving one. IBM translates computer-related documents into no less than 23 languages. Moreover, the volume of documents to be translated is growing dramatically. Therefore, improving translator productivity through the increased mechanization of the translation process is an important business goal.

The IBM Research Division is actively involved in basic research on natural language processing relevant to both machine translation and machine-aided human translation. This effort includes the development of a number of prototype machine translation systems. Let me briefly describe four of them.

IBM's English-to-Japanese system, called SHALT, which stands for "System for Human Assisted Language Translation," is tailored for the translation of computer-related documents. The SHALT system is embedded in the IBM-developed Translation Support System—TSS—which provides the word processing tools necessary for efficient post-editing of translated material.

SHALT is currently being used internally for translating computer manuals from English into Japanese at the IBM Japan translation center. On the technical side, SHALT is a transfer-based system that includes a broad coverage English grammar for analysis, a transfer component involving over 500 transformational rules relating English and Japanese sentence structures, and a large Japanese grammar for generating Japanese sentences.

IBM's Japanese-to-English translation system, JETS, is a research prototype that is not tailored to any particular subject area. Rather, the goal of the JETS project is to produce a core machine translation system that can be easily customized for special purposes. JETS is also a transfer-based system.

One noteworthy feature of JETS is a Japanese workbench that supports interactive analysis of Japanese source sentences. That is to say, the Japanese analysis procedure prompts a user to provide information in cases where there is ambiguity that the system cannot resolve.

IBM's Logic-based Machine Translation system—LMT—is an experimental shell system that has been developed in the framework of logic programming. Versions of LTM exist in various stages of development for translation from English to German, Danish, Spanish and Hebrew, as well as from German and Danish to English.

The analysis component uses a grammar formalism called Slot Grammar, a particular type of logic grammar developed at IBM, in which much of the information used in analysis is contained in the dictionary.

One major emphasis of this work has been on achieving generality of procedures as well as modularity and extensibility of the system as a whole. This design philosophy makes it comparatively easy to add language pairs.

In addition to the above efforts, IBM has recently begun research on a statistical approach to machine translation from French to English. The basic vocabulary comes from the bilingual proceedings of the Canadian Parliament as recorded since 1973. The novelty of this approach is the integration of different sources of information into a uniform probabilistic framework and the use of a very large body of data—3 million pairs of sentences—from which the parameters of a detailed statistical model of the translation process are estimated.

A number of researchers at IBM foresee integrating statistical approaches to machine translation with more traditional, linguistics-based approaches.

Another area of focus at IBM Research is on lexical systems, that is, the design, development and maintenance of very large, computerized dictionaries, terminological databases and so forth, which are suitable for use in advanced translator work stations, as well as in more fully automated machine translation environments.

Topics of relevance here include the automatic extraction from text of words and phrases not already registered in a dictionary, the recording and manipulation of previously translated material for potential use in future translations, and statistics-based identification of bilingual correspondences in large corpora of parallel translations.

While high quality, fully automated translation is not, in our opinion, on the immediate horizon, we do anticipate in the near term advanced translator work stations that provide not only a variety of sophisticated translation tools but also interactive automated translation as one option available to the human translator.

Work toward this goal is already under way at IBM, targeting in-house translation of various European languages. IBM is also conducting two joint projects with universities in Asia to develop English-to-Korean, Korean-to-English, and English-to-Chinese machine translation systems.

Much of the above work has drawn heavily upon the results of long-term research in such areas as theoretical linguistics, computational linguistics, and artificial intelligence, conducted both within IBM and in the academic and technical communities at large.

Continued progress toward the goal of highly automated machine translation is intimately tied to future advances in these supporting fields, fields which have benefited from support of Federal agencies such as the National Science Foundation and DARPA. Continued Federal support of these underlying areas would serve to advance machine translation technology faster than it might otherwise proceed.

I would like to close with some general observations on the status of machine translation technology. In 1966, at the time of the so-called ALPAC report, machine translation was, broadly speaking, thought of as a batch operation primarily dealing with already printed, technical documents, and the chief purpose of ma-

chine translation was information gathering. The technological environment for machine translation was, by today's standards, primitive.

Over the intervening period, the tremendous advances in the computer industry have created a much more flexible and interactive computer environment. Part of this continuing trend will be the development of highly flexible, interactive translation systems helping to meet a variety of information processing needs spanning the spectrum from passive information gathering to the active creation of documents and their dissemination in many languages.

As we strive to mechanize the translation process, it is important to bear in mind that the concept of machine translation is multifaceted. In the area of translation, there is not, practically speaking, one solution to one well-defined translation problem.

For this reason, we at IBM take the view that work in this area should be balanced, focusing both on tools for machine-aided translation and on highly or even fully automated machine translation systems.

Thank you, Mr. Chairman.

[The prepared statement of Dr. Johnson follows:]

STATEMENT OF DAVID E. JOHNSON
RESEARCH SCIENTIST, IBM CORPORATION
THOMAS J. WATSON RESEARCH LABORATORIES
BEFORE THE
SUBCOMMITTEE ON SCIENCE, RESEARCH AND TECHNOLOGY
UNITED STATES HOUSE OF REPRESENTATIVES
SEPTEMBER 11, 1990

Good morning, Mr. Chairman, and members of the subcommittee. My name is David Johnson. I am a research scientist at IBM's T. J. Watson Research Laboratories in Yorktown Heights, N.Y., where I am involved in research on machine translation. This morning, I will provide a summary of IBM's work in this important area and, of course, will be glad to answer any questions you might have.

Our interest in machine translation is, in part, a self-serving one. IBM translates computer related documents into no less than 23 languages. Moreover, the volume of documents to be translated is growing dramatically. Therefore, improving translator productivity through the increased mechanization of the translation process is an important business goal. The IBM Research Division is actively involved in basic research on natural language processing relevant to both machine translation and machine-aided human translation. This effort includes the development of a number of prototype machine translation systems. Let me briefly describe four of them.

IBM's English-to-Japanese system, called SHALT, which stands for "System for Human Assisted Language Translation", is tailored for the translation of computer-related documents. The SHALT system

is embedded in the IBM-developed Translation Support System (TSS), which provides the word-processing tools necessary for efficient post-editing of translated material. SHALT is currently being used internally for translating computer manuals from English into Japanese at the IBM Japan translation center. On the technical side, SHALT is a transfer-based system that includes a broad coverage English grammar for analysis, a transfer component involving over 500 transformational rules relating English and Japanese sentence structures, and a large Japanese grammar for generating Japanese sentences.

IBM's Japanese-to-English translation system, JETS, is a research prototype that is not tailored to any particular subject area. Rather the goal of the JETS project is to produce a core machine translation system that can be easily customized for special purposes. JETS is also a transfer-based system. One noteworthy feature of JETS is a Japanese workbench that supports interactive analysis of Japanese source sentences. That is to say, the Japanese analysis procedure prompts a user to provide information in cases where there is ambiguity that the system cannot resolve.

IBM's "Logic-based Machine Translation" system (LMT), is an experimental shell system that has been developed in the framework of logic programming. Versions of LMT exist in various stages of development for translation from English to German, Danish, Spanish and Hebrew, as well as from German and Danish to English. The analysis component uses a grammar formalism called

Slot Grammar, a particular type of logic grammar developed at IBM, in which much of the information used in analysis is contained in the dictionary. One major emphasis of this work has been on achieving generality of procedures as well as modularity and extensibility of the system as a whole. This design philosophy makes it comparatively easy to add language pairs.

In addition to the above efforts, IBM has recently begun research on a statistical approach to machine translation from French to English. The basic vocabulary comes from the bilingual proceedings of the Canadian Parliament as recorded since 1973. The novelty of this approach is the integration of different sources of information into a uniform probabilistic framework and the use of a large body of data (3,000,000 pairs of sentences) from which the parameters of a detailed statistical model of the translation process are estimated. A number of researchers at IBM foresee integrating statistical approaches to machine translation with more traditional, linguistics-based approaches.

Another area of focus at IBM Research is on lexical systems, that is, the design, development and maintenance of very large, computerized dictionaries, terminological databases and so forth, which are suitable for use in advanced translator workstations, as well as in more fully automated machine translation environments. Topics of relevance here include the automatic extraction from text of words and phrases not already registered in a dictionary, the recording and manipulation of previously

translated text for potential use in future translations, and statistics-based identification of bilingual correspondences in large corpora of parallel translations.

While high quality, fully automated translation is not, in our opinion, on the immediate horizon, we do anticipate in the near term advanced translator workstations that provide not only a variety of sophisticated translation tools but also interactive automated translation as one option available to the human translator. Work toward this goal is already underway at IBM, targeting in-house translation of various European languages. IBM is also conducting two joint projects with universities in Asia to develop English-to-Korean, Korean-to-English and English-to-Chinese machine translation systems.

Much of the above work has drawn heavily upon the results of long-term research in such areas as theoretical linguistics, computational linguistics, and artificial intelligence, conducted both within IBM and in the academic and technical communities at large. Continued progress toward the goal of highly automated machine translation is intimately tied to future advances in these supporting fields, fields which have benefited from support of federal agencies such as the National Science Foundation and DARPA. Continued federal support of these underlying areas would serve to advance machine translation technology faster than it might otherwise proceed.

I would like to close with some general observations on the status of machine translation technology. In 1966, at the time of the so-called ALPAC report, machine translation was, broadly speaking, thought of as a batch operation primarily dealing with already printed, technical documents, and the chief purpose of machine translation was information gathering. The technological environment for machine translation was, by today's standards, primitive. Over the intervening period, the tremendous advances in the computer industry have created a much more flexible and interactive computer environment. Part of this continuing trend will be the development of highly flexible, interactive translation systems helping to meet a variety of information processing needs spanning the spectrum from passive information gathering to the active creation of documents and their dissemination in many languages.

As we strive to mechanize the translation process, it is important to bear in mind that the concept of machine translation is multi-faceted. In the area of translation, there is not, practically speaking, one solution to one well-defined translation problem. For this reason, we at IBM take the view that work in this area should be balanced, focusing both on tools for machine-aided human translation and on highly or even fully automated machine translation systems.

Thank you, Mr. Chairman.

Mr. VALENTINE. Thank you.

Mr. Zaretsky?

Mr. ZARETSKY. Thank you, Mr. Chairman. I hope you can hear me, I don't see the light on my microphone.

I would like to say that I consider it a privilege to have the opportunity to appear before the Subcommittee. I was invited here to present the businessman's point of view. I, therefore, would like to talk about machine translation from the standpoint of both, of my company—a technical translation company—and from the standpoint of my clients—corporations operating in the global marketplace.

I'd like to start with a brief case study. One of my clients—a heavy equipment manufacturer—recently lost a major contract in North Africa to a Japanese competitor. My client's equipment is world-class. The service and other conditions of the proposed contract were also excellent. Yet they lost it, and here is why:

The contract required all training materials and also operating and maintenance manuals to be supplied in the local language, and in that case French. Due to the variety of equipment to be supplied under this contract, total documentation amounted to about 7,000 pages. In the translation industry, 7,000 pages represents about 3-1/2 man-years of work.

The duration of the entire project in North Africa was supposed to be 18 months. Moreover, some of the training documentation under the terms of the contract were supposed to be supplied only three months after the contract would be signed.

My company is capable of delivering the material in French in the time required, but with the deadline it would cost our client \$1.2 million. That is three times more than we would normally charge.

The Japanese already have their material in French. So they got the contract.

You might think that my client should have known better and had its material translated just as the Japanese did. Well, you, of course, would be at least partially right. We Americans are still guilty of believing that English is the language of international trade. It is—on a certain level. But not when it comes to operating a processing line in Turkey or providing a maintenance to a bulldozer in the gold mines of Siberia. The Japanese have a little psychological advantage over us—they never expect anybody to speak a little bit of Japanese.

The fact remains that translation is still a very labor-intensive process. Not much has changed since the days when monks sat over scripts and translated it with quill in the hand. The tools, perhaps, change. Maybe instead of quill, we have word processors today. But it is still—a man looking up the words in a dictionary and using this—the most sophisticated computer we have to do the translation process.

Obviously the question arises, why doesn't the translation industry today, of which I am a part, use machine translation?

Well, primarily because existing automated translation systems today are based on the technology of the '60s and '70s. And we all know what dramatic technological change took place in the past

10, 20 years. There are no user-friendly systems on the market which utilize state-of-the-art technology.

Alpnet Corporation, which started out selling machine translation, eventually finds itself in the translation business, providing translation services, because there are no takers for their very complex, very cumbersome \$60,000 machine translation, machine-assisted translation system.

Their major competitor, WCC—also known as Weidner Corporation—was acquired by Bravice International, a Japanese company. And before they were acquired they also found themselves providing translation services because the machine-assisted translation software was too complex to use and produced too mediocre results.

Systran—of which we've heard so much today—one of the early translation software developers, now leases its programs for \$3,000 a month, or allows users to top its essential data banks for a fee per word. It has about 20 clients.

These systems must still be used by professional translators who have to go back over the text with a fine-toothed comb searching for errors. The marketplace has simply judged that these systems are not cost-effective.

My company would be among the first to use these systems if they were good. We've tried all the commercially available machine-aided translation systems—and given up on them, because they are too cumbersome and unreliable.

Since we don't have the capability to come up with our own system, we've resorted to developing a set of translator's tools to make our work easier—for example, a Russian language word processor, spellcheckers in various foreign languages, specialized on-line glossaries for the clients' specific technologies, et cetera. We don't have the time or resources to do more. I was delighted to hear that at IBM—Dr. Johnson—is developing some of the tools to ours.

This, then, is the state of the industry in which worldwide sales for 1988 are estimated to be \$20 billion—\$20 billion. Yet, a major exporter such as J.I. Case Company or Dresser Industries translates only 3 to 5 percent of what they would like to translate—only 3 to 5 percent—because it is still too expensive and too time-consuming.

Machine translation is not the all-purpose answer to every translation need. Just as there are many different kinds of computers, there should be many different forms of MT for various translation needs. Machine translation can be highly efficient for producing large volumes of technical material in which terminology can be standardized—such as the manuals that accompany the equipment and technology the American companies want to export.

But these solutions cannot be found without government assistance. Government policies can help increase export sales by encouraging the development of this technology. Machine translation can be a cost-effective tool that will enable more companies to sell goods and services overseas and thus compete more successfully in the global marketplace.

The Japanese are now funding a considerable amount of the most advanced research being done on machine translation in this country. U.S. Government funding will help ensure that this tech-

nology remains in this country, keeping us at least on par with our major global competitor.

It will also foster the development of a modern translation industry here in the United States. In addition to aiding our own companies in their efforts to sell abroad, a highly efficient translation industry in this country has the potential to attract translation business away from Europe and Japan—from American and foreign companies—and then the United States will be a viable marketplace.

My company currently employs a 150 free-lance translators. I take pride in the fact that so many people make a living here. Access to efficient machine translation will make us even more competitive and efficient. It will permit us to employ even more people.

There are over 3,000 members in the American Translators Association today, and the ranks are constantly growing. Only 10 years ago the same Association had only 1700 members. This is just another sign that the world is becoming one large marketplace.

Who can tell what industry will experience the most dynamic growth in the 21st century?

In this age of information it very well may be the newborn industry of technical translation.

I believe that we need to direct our efforts to develop sound principles on which to base machine translation systems. Experience has proven that bringing less-than-perfect systems to the market has only made potential users skeptical of the entire concept of machine translation. By supporting the efforts of institutions such as Carnegie Mellon university, the Federal Government can lay the groundwork for our technological superiority in this field.

Thank you.

[The prepared statement of Mr. Zaretsky follows:]

the corporate word, inc.

...the language management company

A Businessman's Perspective on Machine Translation

Gregory Zaretsky, President

I consider it a privilege to have the opportunity to appear before the Subcommittee. I was invited here to give a businessman's opinion on machine translation, and I would like to talk about it from the standpoint of both my own company, a translation company, and from the standpoint of my clients, corporations operating in the global marketplace. I'd like to start with a brief case study.

One of my clients (a heavy construction equipment manufacturer) recently lost a major contract in North Africa to a Japanese competitor. My client's equipment is world-class. The service and other conditions of the proposed contract were also excellent. Yet they lost it. Here's why.

The contract required all training materials and operation and maintenance manuals to be in the local language, French. Due to the variety of equipment to be supplied under this contract, total documentation amounted to 7,000 pages. In the translation industry, this represents 3.5 man years of work. The duration of the entire project in North Africa was supposed to be 18 months. Moreover, most of the documentation such as operators' manuals and overhead slides for training seminars were needed as soon as three months after the contract was signed.

My company is capable of delivering the material in French in the time required, but with that deadline it would have cost our client \$1.2 million. That is three times more than normal.

The Japanese already had their material in French. So they got the contract.

You might think that my client should have known better and had its material translated just as the Japanese did. You of course would be at least partially right. We Americans are still guilty of believing that English is the language of international trade. It is -- on a certain level. But not when it comes to operating a processing line at a food cannery in Turkey or repairing a bulldozer in the gold fields of Siberia. The Japanese have a psychological advantage over us -- they never expect anybody to speak even a little Japanese.

The fact remains that translation is still a very labor-intensive process. Not much has changed from the days when monks sat over ancient manuscripts translating them with quill pen in hand. By and large, translators, whether they work directly in corporations or work out of their homes for translation agencies, still sit with dictionaries, looking up words. At most they have a word processor and perhaps a few primitive on-line dictionaries. A skilled, conscientious translator produces an average of 8 double-spaced pages of text a day. Why doesn't he use machine-assisted translation?

the corporate word, inc.

Existing automated translation systems are based on the technology of the 60's and 70's. There are no user-friendly systems on the market which utilize state-of-the-art technology. Alpnat Corporation, which started out selling machine-assisted translation software, had to become a translation company because there were no takers for its very complex \$60,000 system which produces mediocre results. Their major competitor, WCC was acquired by Bravice International, a Japanese company. Systran, one of the early translation software developers, now leases its programs for \$3,000 a month or allows users to tap its central data banks for a fee per word. It has about 20 clients. These systems must still be used by professional translators who have to go back over the text with a fine-toothed comb searching for errors. The marketplace has simply judged that these systems are not cost-effective.

My company would be among the first to use these systems if they were good. We've tried all the commercially available machine-aided-translation systems and given up on them because they are too cumbersome and unreliable. Since we don't have the capability to come up with our own system, we've resorted to developing a set of "translator's tools" to make our work easier -- for example a Russian-language word processor, spell-checkers in various foreign languages, specialized on-line glossaries for our clients' technologies, etc. We don't have the time or resources to do more.

This, then, is the state of an industry in which worldwide sales for 1988 are estimated to be \$20 billion. Twenty billion dollars -- yet major exporters like J.I. Case Company or Dresser Industries translate only 3-5% of what they would like to translate because it is still too expensive and time-consuming.

Machine translation is not the all-purpose answer to every translation need. Just as there are many different kinds of computers, there should be many different forms of MT for various translation needs. Machine translation can be highly efficient for producing large volumes of technical material in which terminology can be standardized -- such as the manuals that accompany the equipment and technology the American companies want to export.

But these solutions cannot be found without government assistance. Government policies can help increase export sales by encouraging the development of this technology. Machine translation can be a cost-effective tool that will enable more companies to sell goods and services overseas and thus compete more successfully in the global marketplace.

The Japanese are now funding a considerable amount of the most advanced research being done on machine translation in this country. U.S. government funding will help ensure that this technology remains in this country, keeping us at least on par with our major global competitor. It will also foster the development of a modern translation industry here. In addition to aiding our own companies in their efforts to sell abroad, a highly efficient translation industry in this country has the potential to attract translation business away from Europe and Japan -- both from American and foreign companies. My company currently employs 150 free-lance translators. I take pride in the fact that so many people make a living here. Access to efficient machine translation will make us even more competitive and efficient. It will permit us to employ even more people. There are over 3000 members in the American Translators Association today, and the ranks are constantly growing. Only ten years ago the same association had only 1700 members. This is just another

the corporate word, inc.

sign that the world is becoming one large marketplace. Who can tell what industry will experience the most dynamic growth in the 21st century? In this age of information it very well may be the newborn industry of technical translation.

I believe that we need to direct our efforts and develop sound principles on which to base machine translation systems. Experience has proven that bringing less-than-perfect systems to the market has only made potential users skeptical of the entire concept of machine translation. By supporting the efforts of institutions, such as Carnegie Mellon University, the Federal government can lay the groundwork for our technological superiority in this field.

#

GREGORY ZARETSKY
President and CEO

Employment

- 1979- The Corporate Word, 18 South, Gateway Center 3, Pittsburgh, PA
President
In addition to general management responsibilities, conceptualize requirements for machine-assisted translation systems, including foreign-language word processing and typesetting systems, Arabic and other non-Latin fonts, and relevant software application packages.
- 1973-1979 Pullman Incorporated
- 1978-1979 Manager of special projects reporting to VP and general manager, Pullman Swindell International Company.
Developed and carried out barter transaction related to company projects abroad: headed trade negotiations with representatives of the People's Republic of China, PRC, Poland, Hungary: created independent entity responsible for worldwide marketing of goods received in barter.
- 1976-1978 Manager, International Sales reporting to Senior VP, Pullman, Inc.
Advised US companies desiring to enter new foreign markets on specific approaches to those markets; acted as support to company foreign offices and liaison between headquarters, foreign offices, and domestic and foreign clients; represented company with foreign delegations to further existing and developed new sales opportunities.
- 1973-1976 Manager, Translation Services reporting to VP, Administration, Pullman Swindell
Reorganized translation services from support unit to profit center; developed market strategy; Managed directly 30 full-time and 50 free-lance translators responsible for translation of critical technical, legal, and commercial documents; increased department's profits 140% over 1 year.
- 1972-1973 University of Pennsylvania, Philadelphia, PA
Asst. Professor of Russian Language and Literature,
Department of Political Science and Foreign Affairs.
- 1969-1972 Professional Photographer, Philadelphia, PA
Specialized in commercial and industrial photography

Education

- 1974-1976 Management training courses at the University of Pittsburgh, Wharton School, General Electric Institute
- 1976 MA, International Studies, University of Pennsylvania
- 1969 Fellowship, Russian Studies, Moscow State University
- 1968 BA, Russian Studies, Jagiellonski University, Cracow, Poland

Mr. VALENTINE. Thank you, sir. Thank all of you.

I would like very much to have some information which I'm not going to ask you for, but, you know, we have to deal with. We have to try to classify things in this Committee and in this country, and one not related with this—with the developments in this area, this very fascinating and important area, might wonder where this—the government's serious participation in R&D efforts would—where it would rank on the scale of things that we have to do.

One in my position maybe would naturally think of the awful expense and risk to this Nation at this particular time when almost paralleling the time that Congress left here for the August work period, we have developed the problems—or the problems that developed in the Persian Gulf. And there we stand spending money and exposing our troops to supply the lifeblood of freedom to the rest of the world, that a lot of countries don't seem to be carrying their full burden of responsibility—Japan, in particular, that seems to eat us up at every—on every turn. Maybe well they—well they can afford to, as long as we continue to do all these things to ensure their well-being.

Aren't we talking here about a field that is somewhat different from a more pressing national interest? Most of the scientific developments that we have considered in this subcommittee and full committee come down to enhancement of a quality of life in this country, or national defense, or better television sets, or something of that kind.

Aren't we here talking about the—purely and simply about the most convenient way to translate a book from one language to another, and the questions are how they get the most speed and the greatest accuracy? And does it really make that much difference who does it?

I know there's a great market out there, but if Japan is ahead—Japan is ahead of us, is it in our national interest to try to duplicate what they do? Why don't we just go buy this stuff from the Japanese? Aren't we really talking about language tools?

Dr. Johnson?

Dr. JOHNSON. I think some earlier remarks made today were relevant to this point, Mr. Chairman, namely, that the needs of U.S. users is perceived to be different than the Japanese who are—have been using—attempting to use machine translation for information and dissemination, that is product documentation, and have some control over the source text.

Whereas, the—one of the major needs in the United States—this is not IBM's personal need. IBM's Corporation's need is for information acquisition, that is, the perception is that there's a large body of valuable technical information in Japanese and that there is not a skilled pool of translators available to meet translation needs of people in the technical community. And, therefore, that there is a certain amount of vital information not being made available to people in this—researchers in this country.

Mr. VALENTINE. Well, if the Japanese are—are the Japanese ahead of us in this area? And—I assumed that a Japanese machine that would translate English-into-Japanese was like a car—you could put it in reverse—just change gears and do the reverse.

It's not that simple, is it?

Dr. JOHNSON. Well, no, it's not that simple. But I think there's another point, too, that I'd like to address, which is that there are systems in Japan, but in my estimation and the estimation of many of the people working in the area, they're not very good. And so—now here I assume you're talking about automated machine translation as opposed to translators tools. And if we look at that area, in my view, we find that the state of the art is not such that we have good quality systems.

It's true that the Japanese are ahead in actually developing systems. It's not clear that they have been successful commercially—I believe they have not. And probably—and I believe for good reason.

So then I believe the question is—that we must address—is the need for translation of primarily Japanese into English in this country. And then the next question is to assess what the state of the art is and what potential for mid-term—well, near-term and then mid-term systems would be—and not rely on the Japanese for leadership in this area.

One area where we might cooperate with them and where we would hopefully be able to learn from them is in the area of analyzing the Japanese language. But beyond that, most of the technology in their systems in fact goes back to the scientific and technological base that had been developed in this country 10, 20 years ago in fields such as linguistics and artificial intelligence.

Mr. VALENTINE. Isn't this an area which would seem to lend itself to international cooperation to a much greater extent than most situations that we find ourselves in because we're not dealing, are we, with national security questions, with matters of great importance as far as the disclosure of our secrets? We're just trying to find out a better way to translate all those characters into A, B, C, and so on.

Dr. JOHNSON. Yes, Mr. Chairman, I believe that's quite correct.

Dr. CARBONELL. May I add?

Mr. VALENTINE. Yes, sir.

Dr. CARBONELL. Excuse me, Mr. Chairman, may I add a remark to your original question—

Mr. VALENTINE. Yes, sir.

Dr. CARBONELL. —about the significance of mission translation? And I believe that the significance is perhaps indirect but, nonetheless, extremely important as it relates to our international competitiveness in trying to market our products in the world economy.

The example that Mr. Zaretsky quoted about the heavy equipment manufacturer losing a major contract in a North African country because of the lack of speedy translation is something that I've observed repeatedly, multiple times across a wide spectrum of our industry.

One of our greatest national problems right now is the international trade imbalance that we're suffering, especially with Japan; which would be greatly improved if we could automatically, quickly and reliably translate documentation on marketing goods and everything else into the target country into which we were selling.

Mr. VALENTINE. Well, are you saying that the Japanese were able to do that because they took the time to perfect the translations, or they had the equipment that would help them to do it better than we did? Certainly we—an American company should

have anticipated that, as was mentioned, that there was—there might be this need.

Dr. CARBONELL. Of course, the Japanese have a much, much larger group of trained translators into English and into French, and so forth, than we have trained for translation into Japanese, or Korean, or Chinese, and so forth.

So one answer would be to start training 20,000 people and become translators and keep them employed. Another answer is to try to automate the process of machine translation, or perhaps a combination of the two answers is important. But we're losing potential global market because of our limited linguistic capabilities.

Mr. VALENTINE. Well, I guess—what concerns me on that was that we hear—I think we heard at the beginning of the testimony—that this is another area where the Japanese are way ahead of us.

And, you know, if that's the case, it seems to me that if this is an area where there is no acute dynamic national interest involved. If we're just talking about translations, then what is the answer to those people that say, we don't need to spend our money, we just go buy what they got.

I know it would be great for American industry to have the \$20 billion that this new industry would attract.

Well, let me just stop there and ask you to comment on that wandering question.

Dr. CARBONELL. What they've got, or what they're striving to have, since their work is not yet complete in translation—we've heard their accuracy is not yet up to what they desire—is ways of accurately translating Japanese manuals into English.

We do not want to translate Japanese manuals into English; we want to translate English manuals into Japanese, and Korean, and Chinese, and German, and French, and Spanish. We have a different task. Perhaps starting from what the Japanese have, one may be able to build on top of it and adapt it, or starting on what we already have here, or a combination of the two.

But the needs are not the same. We don't want to help the Japanese market their products in our country. They already do a damn good job of that. We want to go the other way around.

Mr. VALENTINE. Changing—well, changing the subject—what does the U.S. Task Force Study Team, which you will be heading this Fall, which is sponsored by NSF, the Department of Commerce and the Defense Advanced Products Research, or DARPA, what does it hope to accomplish in Japan?

Dr. CARBONELL. Well, there are the stated objectives of the JTEC process, and there are some additional objectives we also wish to accomplish—those are the JTEC processes are primarily an assessment of the state of the art of the technology in the research laboratories and in the pre-competitive product development to know what's coming in the immediate future as well as what is being cooked for the longer term future.

We're trying to assess the capability of the Japanese to deploy productive use for large scale machine translation and assess it, not just from its impact on the marketplace and the scientific principles, but also to determine how one may be able to exploit their

capability, where it is well developed, for the American translation effort as well. That's the official agenda, write up this assessment.

Unofficially, we also will be developing a set of recommendations as to what to do next from this—from what we learn in Japan and from what we already know is occurring elsewhere. And these recommendations are one of the ingredients that I hoped would go towards building this longer term plan of how we get from here to there that was alluded to by Mr. Walgren, that I would very much like to see in place.

To state it bluntly, we have no such plan now. We are in the initial process of gathering data in order to be able to build such a plan. I would very much like to see one.

Mr. VALENTINE. What are the primary technological advances in machine translation in recent years? What is it that gets us away from what the code-breakers in World War II discovered?

Dr. Bennett?

Dr. BENNETT. I think of the kinds of things that we've made advances in—linguistic theory and computation linguistic—the machine area is much better. Just on our own project we've gone from a DEC computer to symbolics list machine, which is state of the art, very fast. I think we're more aware of the realities that machine translation and what we're about. I think up to, and perhaps including the ALPAC era, there was a sense of machine translation was to—you know, there was going to be one approach to handle everything, sort of everything from newspaper articles to poetry. And I think we've become aware of our limitations on that.

I think we have a lot more experience to—this is clearly the case just in our own project. I didn't cite METAL simply to cite METAL, but rather to indicate that we have considerable experience just at the University of Texas and this is not—we have considerable experience with other projects and other places in the area of machine translation, and this kind of technology that 30 years counts for something.

Mr. VALENTINE. Anybody else like to add to that, or take from it?

Dr. CARBONELL. I would like to add one remark, to the effect, just to give you an example of the kinds of developments that have been occurring lately.

One kind of—the most dominant kind of translation, as was cited by Dr. Johnson from IBM, is a transfer approach, that is, you start with language in the source text; you do some analysis of that, using a grammar and dictionary. And then you have a set of specific rules—over 500 in the case cited, that convert English structures into Japanese once or—excuse me—in this case it was English into Japanese structures—and then another grammar to generate the Japanese text.

A different paradigm that is starting to be quite successful is called the interlingual approach. Where instead of having these transfer rules that depend on both language as source and target, you analyze the source language for meaning; that is, you understand what it's about. You represent this in an internal form, a logical form in the computer. And once you've done that, you can generate it in any target language you want.

This is very important for international marketing if you have to translate a manual into 23 different languages—you analyze it

only once. You don't have to transfer it 23 times and develop 23 times that number of large rules. And this is an advance that didn't exist back then.

It is just an example of our technical advance. There are many others that together will contribute to making machine translation far more effective than back then. It is difficult to state them in detail without starting to draw diagrams and so forth.

Mr. VALENTINE. Are there different scientific schools of thought in machine translation technology?

Dr. CARBONELL. Yes.

Mr. VALENTINE. Okay.

Mr. Walgren?

Mr. WALGREN. Thank you, Mr. Chairman. And I particularly appreciate the last panel because on the one hand, two of the four come from Pittsburgh, and I—I like to think that the French philosopher Serban Schrifer once said that Pittsburgh is where the world is being reinvented. And you folks are very good examples of what is happening there.

But also because of how positive a panel you are and how you underscore the potential that machine translation has and the importance that it has for the wide range of economic activity in a world of—in a world economy.

I had—we had focused earlier on the Japan to English translation, and much of our political energy is directed in that area. But Dr. Carbonell goes much beyond that in wanting to—in focusing on the importance of being able to translate into all languages.

The Task Force that you'll lead, will it be focused on Japan to English? I understand that the Science and Technology Agreement is sort of focused on that because that's what we were asking, and that's what the Japanese were trying to do probably for reasons that escaped us.

Dr. CARBONELL. The answer is no; it will be focused on Japanese technology. The Japanese are translating into many languages beyond English—the European languages, and more importantly, the Asian languages—Korean, Mandarin, Cantonese, Thai, and so forth—that we do not have that level of expertise here, nor is it existent in the European environment.

So, therefore, we will be analyzing the Japanese capabilities into all of these languages, not just Japanese-English.

Mr. WALGREN. [presiding] I see.

I don't know whether we got earlier the actual mission statement that you received. And I guess we ought to ask it from you, and if you could get us that in a submission form of some kind. There must be some written document that you're given.

Dr. CARBONELL. There is a draft of such a document. I would prefer to get the okay from the entire panel prior to submitting it and then I will be delighted to submit it.

Mr. WALGREN. Sure.

[The document follows:]



UNITED STATES DEPARTMENT OF COMMERCE
 National Technical Information Service
 5285 Port Royal Road
 Springfield, Virginia 22161
 OFFICE OF THE DIRECTOR

U.S.-Japan Agreement on Cooperation in Science and Technology

Task Force on Scientific and Technical Information

A key objective of the Task Force is to increase the translation of scientific and technical information.

The Department of Commerce, Department of Defense and the National Science Foundation are organizing a team of experts to assess the potential of computer technology to help achieve this goal.

The primary objective of the "machine translation study team" is to assess the state-of-the art of machine translation research, development and deployment in Japan, focusing on research and development efforts targeted at producing complete deployable systems in the foreseeable future (within 7 years). Presently deployed systems with users performing actual translations are also of extremely significant interest, especially with respect to evaluation, user acceptance, and plans for continued system development.

Secondary objectives include:

- Studying recent development in machine aids to human translation (MAT).
- Focusing on the scientific underpinnings of present and future machine translation systems (parsing algorithms, linguistic formalisms, and so on).
- Examining scaling-up properties as machine translation laboratory prototypes transition to deployed systems (management of ~~dictionary and~~ terminology banks, attaining full grammatical coverage, improving quality of translation, etc.)
- Exploring new approaches to machine translation (knowledge-based machine translation, translation of discourse as well as text, speech-to-speech translation, unification grammars, and so forth).
- Understanding the dynamics of research and development organizations that develop successful machine translation systems.
- Exploring possibilities for collaboration with American laboratories.

Mr. WALGREN. But then you also went beyond that to say that you hoped the panel will be able to make recommendations about where we go from here. Is that part of the charge as well, or is it more part of the expectation?

Dr. CARBONELL. That is part of the understanding, I suppose that maps into your expectation. It is not part of the official JTEC charge because it's outside the agreement that was signed between the head of states of the two countries to make complete recommendations.

We will, nonetheless, make them outside of the JTEC report itself.

Mr. WALGREN. Well, we would like to encourage you in that, and not to create any problems with the structure in which you're going to be working on the formal side of that report. But I hope all your panel will know that this Committee is very interested in the direction that you can give us as well as recommendations you can make directly to the Federal Government.

And as you underscore, we have no plan. Several of us have been asking for a plan for a long period of time and I hope that your work might directly lead us to be able to cobble one together, if it is not formally part of the work product.

Dr. CARBONELL. Thank you.

May I encourage you to suggest a target date by which a plan may be proposed, or any other kind of forcing function to get us to actually do it?

Mr. WALGREN. Well, you're going to be reporting in the spring, as I understand it. Is there a date for that at this point?

Dr. CARBONELL. There's no firm date yet.

Mr. WALGREN. Well, I would hope that if this Committee would formally ask you, and we will, for not only your report to the Department of Commerce but your views and suggestions as to what work needs to be done, and how you would get from here to there, what it would take. And if you could—we'll get you a formal request for that, and I would think perhaps one month after the plan—the actual assessment report is made to the Department of Commerce—this Committee would like to have a road map of where your group suggests we should go, and what—some relevant time frames or other targets might be for anyone planning a successful pursuit of this area.

Dr. CARBONELL. Sounds good.

Mr. VALENTINE. All right.

Well, on behalf of Mr. Valentine, who had to go, let me express our appreciation to all of you. We consider you not only a valuable resource to the Committee but a tremendous resource to the Country and we are going to see this through successfully somehow.

Thanks very much.

[Whereupon, at 12:35 p.m. the subcommittee was adjourned.]

APPENDIX

Report of a Symposium on
Japanese to English
Machine Translation

Held at the National Academy of Sciences
in Washington, D.C.
on December 7, 1989

Office of Japan Affairs
Computer Science and Technology Board
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1990

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competencies and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences of the United States is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering of the United States was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievement of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Research Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Research Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice-chairman, respectively, of the National Research Council.

Available from:
Office of Japan Affairs
National Research Council
2101 Constitution Avenue N.W.
Washington, DC 20418

Printed in the United States of America

The Japanese to English translation on the cover is unedited machine translation output provided by Systran.

STEERING COMMITTEE

Symposium on Japanese to English Machine Translation

Dr. Roger Levien, Chairman
Vice President, Corporate Strategy Office
Xerox Corporation

Dr. Jaime Carbonell
Computer Science Professor
Director of Machine Translation Center
Carnegie Mellon University

Dr. Charles Freiman
Director
The Engineering Foundation

Dr. Richard Samuels
Associate Professor
Department of Political Science
Massachusetts Institute of Technology

PROGRAM
SYMPOSIUM ON JAPANESE TO ENGLISH
MACHINE TRANSLATION

National Academy of Sciences, Lecture Room

December 7, 1989

8:15 - 8:45 AM

Registration

8:45 AM

Chairman's Opening Remarks: Roger Levien,
Xerox Corporation

8:50 AM

Welcome: Lee W. Mercer, Deputy Under
Secretary for Technology, United States
Department of Commerce

9:00 - 10:30 AM

Panel on the State of the Art

Presenters:

Makoto Nagao, Kyoto University
Jaime Carbonell, Carnegie Mellon University

Commentators:

David Johnson, IBM
Alvin Despain, University of
Southern California

10:30 - 12:00 PM

Panel on Market Prospects

Presenters:

Chuck Walrad, Systran
Takehiko Yamamoto, Bravice
International, Inc.

Commentators:

Tom Seal, ALPNET
Cheryl Bettels, DEC-Geneva

12:00 - 12:20 PM

Remarks: Representative George E. Brown Jr.,
United States Congress

12:30 - 1:30 PM

Lunch

1:30 - 3:00 PM

Panel on User Needs

Presenters:

Mark Eaton, MCC
Maria Russo, Xerox Corporation

Commentators:

Dale Bostad, Wright Patterson AFB
Alan Melby, Brigham Young
University/LinguaTech

3:00 - 4:30 PM

Panel on R&D Policy

Presenters:

Bernard E. Scott, Logos
Muriel Vasconcellos, Pan American
Health Organization

Commentators:

Richard Samuels, MIT
Ralph Quinn, Bell Labs
James Unger, University of Hawaii

4:30 - 5:30 PM

(Great Hall) *Informal Reception and
Demonstrations of Machine Translation
Systems and Machine Aids for Translators*

COMMITTEE ON JAPAN

The Committee on Japan has been established to advise the Office of Japan Affairs on its programs, and to assist in defining the contribution that the Academies can make in enhancing U.S. interests through science and technology exchange with Japan.

Harold Brown, Chairman
Johns Hopkins Foreign Policy Institute

Daniel Okimoto, Vice Chairman
Stanford University

Justin Bloom
Technology International, Inc.

Lewis Branscomb
Harvard University

Mac Destler
University of Maryland

Ellen Frost
United Technologies Corporation

Lester Krogh
3M Company

James Merz
University of California,
Santa Barbara

Yoshio Nishi
Hewlett-Packard Company

Terutomo Ozawa
Colorado State University

Ex Officio Members:

Gerald Dinneen, Foreign Secretary, National Academy of Engineering
William Gordon, Foreign Secretary, National Academy of Sciences

Susan Pharr
U.S.-Japan Relations Program,
Reischauer Institute

John D. Rockefeller IV
U.S. Senate

Richard Samuels
MIT-Japan Science and
Technology Program

Roland Schmitt
Rensselaer Polytechnic
Institute

Hubert J. P. Schoemaker
Centecor, Inc.

Ora Smith
Conductus, Inc.

Susumu Tonegawa
Massachusetts Institute of
Technology

OFFICE OF JAPAN AFFAIRS

Since 1985, the National Academy of Sciences and the National Academy of Engineering have engaged in a series of high-level discussions on advanced technology and the international environment with a counterpart group of Japanese scientists, engineers, and industrialists. One outcome of these discussions has been a deepened understanding of the importance of promoting a more balanced two-way flow of people and information between the research and development systems in the two countries. Another result was a broader recognition of the need to address the science and technology policy issues increasingly central to a changing U.S.-Japan relationship. In 1987, the National Research Council, the operating arm of both the National Academy of Sciences and the National Academy of Engineering, authorized first-year funding for a new Office of Japan Affairs (OJA). This newest program element of the Office of International Affairs was formally established in the spring of 1988.

The primary objectives of OJA are to provide a resource to the Academy complex and the broader U.S. science and engineering communities for information on Japanese science and technology; to promote better working relationships between the technical communities in the two countries by developing a process of deepened dialogue on issues of mutual concern; and to address policy issues surrounding a changing U.S.-Japan science and technology relationship.

Staff

Martha Caldwell Harris, Director
Donna J. Audritsh, Research Associate
Karen McDowell, Program Assistant
Karen Duffy, Research Intern

COMPUTER SCIENCE AND TECHNOLOGY BOARD

Joseph F. Traub, Chairman
Columbia University

Alfred V. Aho
AT&T Bell Laboratories

John Seely Brown
Xerox Corporation
Palo Alto Research Center

Michael L. Dertouzos
Massachusetts Institute of
Technology

Samuel H. Fuller
Digital Equipment Corporation

James Freeman Gilbert
University of California,
San Diego

William A. Goddard III
California Institute of
Technology

John L. Hennessy
Stanford University

John E. Hopcroft
Cornell University

Robert E. Kahn
Corporation for National
Research Initiatives

Sidney Karin
San Diego Supercomputer Center

Leonard Kleinrock
University of California,
Los Angeles

Robert Langridge
University of California,
San Francisco

Abraham Peled
IBM T. J. Watson Research Center

Raj Reddy
Carnegie Mellon University

Mary Shaw
Carnegie Mellon University

William J. Spencer
Xerox Corporation

Ivan E. Sutherland
Sutherland, Sproull & Associates

Victor Vyssotsky
Digital Equipment Corporation

Irving Wladawsky-Berger
IBM Corporation

COMPUTER SCIENCE AND TECHNOLOGY BOARD

The Computer Science and Technology Board was established in 1986 to address technical and policy issues in computer science and associated technologies. The Board pursues its mission by promoting the health of the technology and related fields of pure and applied science; initiating studies involving computer science and technology as a critical national resource; responding to requests from the government, nonprofit foundations, and industry for advice on computer science and technology; and providing a base of expertise within the National Research Council in the area of computer science and technology.

The Board's membership, half of which is corporate and half academic, reflects its belief in a partnership of the corporate and academic sectors. In addition, the Board includes a mix of computer and computational scientists, ensuring that the perspective of the computer user is considered.

The Board has an ambitious agenda, focusing on research needs and public policies to enhance U.S. production and use of new computer technologies. Specific priority areas include international competitiveness, high-performance computing, software, talent production, new computer applications, a national infrastructure for future computational technologies, and the articulation of the importance of the field.

Staff

Marjory S. Blumenthal, Executive Director
Damian M. Saccocio, Staff Officer
Herbert Lin, Staff Officer
Margaret A. Knemeyer, Staff Associate
Donna F. Allen, Administrative Secretary
Catherine A. Sparks, Senior Secretary
Pam Rodgers, CSTB Consultant

Contents

PREFACE

1. MACHINE TRANSLATION: FROM A TRANSLATION TO A COMMUNICATIONS AND INFORMATION CHALLENGE.....	1
CHANGING CONTEXT FOR MACHINE TRANSLATION.....	2
SETTING AND MEETING GOALS FOR MACHINE TRANSLATION DEVELOPMENT.....	3
2. THE COMMERCIAL CHALLENGES	8
CURRENT STATUS OF MACHINE TRANSLATION DEVELOPMENT	8
MARKET PROSPECTS	12
USER NEEDS	15
JAPANESE AND U.S. USERS: CONTRASTING NEEDS.....	18
3. THE TECHNICAL CHALLENGES: APPROACHES TO RESEARCH AND ASSESSMENT	20
DEVELOPMENTAL STRATEGIES AND PROBLEMS	21
EVALUATING MACHINE TRANSLATION SYSTEMS.....	25
4. THE RESEARCH AND DEVELOPMENT POLICY AGENDA.....	30
THE SKEPTICS	30
THE ADVOCATES	31
WHAT MIGHT BE DONE?	33
CONCLUSION	36

Preface

At the request of the United States Department of Commerce, the Office of Japan Affairs and the Computer Science and Technology Board of the National Research Council convened a Symposium on Japanese to English Machine Translation on December 7, 1989. With the initiative of a steering committee chaired by Dr. Roger Levien of the Xerox Corporation, the symposium was planned to provide a forum where participants from the private sector and other interested parties could discuss the current state of Japanese to English machine translation, market prospects, and user needs, as well as R&D policy options. Recognizing the potential contribution that machine translation and related technologies can make to the operations of companies doing business around the world, to researchers and policymakers who need to know about developments in Japanese science and technology, and to translators whose work could be augmented by the use of machine aids, the symposium was organized to share information and stimulate thinking about how machine translation and related technologies might address the needs of users in the United States.

This report was prepared by staff members from the Office of Japan Affairs in cooperation with the Computer Science and Technology Board to highlight major points in the symposium presentations and discussions. Participants in the symposium and other experts provided valuable comments and clarifications. This report, however, incorporates a range of views articulated at the symposium rather than a consensus of those presented.

1

Machine Translation: From a Translation to a Communications and Information Challenge

Developmental work on machine translation has been under way for more than 30 years. Some say that we have come a long way, while others question whether the goal is in sight. The reality is that perceptions of what machine translation is—its purpose and scope—are shifting. The paradigm that dominated thinking during most of this 30-year period of research and development (R&D) was the expectation that computer technologies could be developed to build machines capable of doing the work of human translators. Today, machine translation—defined as translation generated by a computer, with or without human intervention—embraces a broad spectrum of technologies. Included are machine translation systems that run on large mainframes and those that run on stand-alone personal computers, enhanced with automatic aids for the human translator.¹ Rather than eliminating human translators, machine translation and related technologies are now seen as ways of facilitating their work.²

The focus of the symposium organized by the Office of Japan Affairs and the Computer Science and Technology Board was machine translation from Japanese to English. Growing interest in machine translation between this language pair

¹ This report deals with computer-based translation, including machine translation systems and machine aids for translators.

² The concept includes machine translation systems that are used primarily for information scanning, where little or no post-editing is done. Such a system can in some circumstances serve as a mechanical translator for a monolingual researcher. Some of the commercial systems available today are beginning to see this kind of usage in limited circumstances.

reflects recognition of the importance of science and technology information produced in Japan. In order to understand the special challenges of Japanese to English machine translation, a comparatively new field of R&D, it is important to appreciate efforts in machine translation involving other languages.

CHANGING CONTEXT FOR MACHINE TRANSLATION

Machine translation technology development has taken on broader significance in an age of rapid international communication and intense market competition. Competition for global markets has intensified the need for companies to get their messages across to overseas customers who speak foreign languages. Companies doing business around the world must be able to speak the languages of their customers. Some large companies have targeted translation technologies as a component of their competitive strategies: IBM sees translation as a "gating" item³ for its marketing objectives in the 1990s; Xerox emphasizes the importance of machine translation in launching products simultaneously in multiple markets. As the importance of global markets grows and cycle times shorten for the introduction of new products, translation increasingly becomes an expensive bottleneck for international companies.

Another, related explanation for changes in perspectives on machine translation is the information explosion. More specifically, U.S. businessmen, researchers and product developers, and policymakers need a better understanding of what is going on in Japan, because that country has now joined the ranks of the world's technological leaders. The need for translations of Japanese technical documents is now apparent, but only a minute fraction of our science and technology community can speak and read Japanese. Growing recognition of the importance of technical information produced in Japan has stimulated interest in the role that machine translation might play in making it possible for Americans to access reports of new inventions, products, and financial developments in Japan.

Changes in thinking about machine translation also reflect the evolution of new concepts of how machine translation systems might be developed and used. Progress in natural language processing technology, the development of more powerful computers, the increasing availability of large, information-laden dictionary data sets, and advances in some aspects of linguistic theory suggest opportunities for R&D. Translations can be delivered through electronic mail and quickly incorporated in successive editions of technical manuals. Interest in "machine-aided translation" has been spurred by the development of workstations

³ Translation of documents is a prerequisite for entering overseas markets.

with dictionaries and other tools that can be used by professional translators. The expectation that translation machines will replace people has now been transformed into the view that these technologies are tools to enhance the efforts of professional translators, researchers, and secretaries.

Today the challenges of machine translation development illustrate the broader challenges of information technology research, development, and use. Machine translation technologies pose a range of theoretical, software, hardware, and even sociological problems that require integration of technologies and improved interactions among developers and users. For all these reasons, machine translation today is more than a linguistics problem. It is a communications and information challenge that demands a diverse range of expertise and resources.

SETTING AND MEETING GOALS FOR MACHINE TRANSLATION DEVELOPMENT

The dream that stimulated early R&D efforts was a machine that produces high-quality translations from a wide variety of source texts at low cost. Even the most ardent supporters of machine translation agree that three decades of effort have not produced the breakthroughs necessary to achieve this dream.

Why the dream remains unfulfilled is the subject of some disagreement. Viewed from one angle, the failure to achieve the goal is a result of giving up too early. Negative evaluations of machine translation in the 1960s were based on the argument that the understanding of text by computer was too difficult, rendering machine translation infeasible.⁴ The ALPAC report by the National Research Council concluded that the basic technology for machine translation had not been developed, and recommended a focus on long-term research in computational linguistics and improvement of translation methods. While the report made no recommendations with regard to funding for research and development on machine translation, the overall negative evaluation of the state-of-the-art is now seen by many as a major cause of the subsequent decline in funding for such research in the United States. Between 1960 and 1970, funding for R&D on machine translation declined from about \$10 million to \$1 million.

⁴ This argument was made by Bar Hillel. The ALPAC report written by the Automatic Language Processing Advisory Committee of the National Research Council, entitled *Language and Machines: Computers in Translation and Linguistics*, is widely seen as the most influential of the studies of machine translation in the 1960s. Published in 1966, the report concluded that the quality of machine translation was poor and cost savings had not been achieved. The report analyzed the products of machine translation at U.S. government agencies after development work had been under way for more than ten years.

Supporters of machine translation say that we would be closer to the dream if we had not given up so soon.⁵

Viewed from another perspective, however, the fact that the initial dream has not been fulfilled is no reason to dismiss the promise of machine translation. The machine translation "heaven" of high-quality, low-cost, general-purpose systems is still distant.⁶ The high-quality systems that exist today are in most cases special-purpose systems working in restricted domains, but none of these are for the translation of Japanese to English. (See Figure 1.) In addition, there are also a number of cost-effective systems that operate in broader domains. Among them, Systran is the only company to have developed a general-purpose commercial system that translates between Japanese and English, as well as 14 other language pairs. A dozen or more prototype machine translation systems have not been able to attain cost effectiveness after a decade of development.

The process of machine translating from Japanese to English and vice versa is comparatively difficult because of important differences in the structures of the two languages. Typical Japanese text consists of Chinese characters and two different styles of Japanese phonetic symbols. These characters and symbols are written without any spaces between individual words, and phrases are rarely separated by punctuation. Grammatically, Japanese differs from English in that it has no distinction of singular and plural nouns, there are no articles, and the subjects are often omitted. In general, the Japanese sentence puts the verb at the end, and the text preceding the verb is in no particular order and contains many compound clauses. The grammatical structure of Japanese may omit pronouns, subjects, and objects, so that the context must be understood in order to choose among alternative possible interpretations. Because of these and many other characteristics of the language, pre-editing is especially helpful to make the text more tractable for machine translation processing.

Most of the Japanese to English systems now in use in Japan succeed because they are limited to particular domains. In the eyes of many, these systems represent a significant step forward, even if they do not fulfill the initial dream.

⁵ In July 1989 the Japan Electronic Industry Development Association published a report entitled: *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, U.S.A.* This report argues that two major conclusions of the ALPAC report are no longer valid: the claim that there is no translation shortage is refuted by estimates of today's translation market in Japan, and numerous examples of successful machine translation are also cited in response to ALPAC's conclusion that it will have no practical use in the near future. The Japan Electronic Industry Development Association's Machine Translation System Research Committee, which prepared the report, was chaired by Dr. Makoto Nagao of Kyoto University and included representatives from Japanese corporations involved in machine translation development.

⁶ Some argue that the major explanation for the failure to reach the target is that there is a much bigger difference between general language and domain-specific language than has heretofore been suspected.

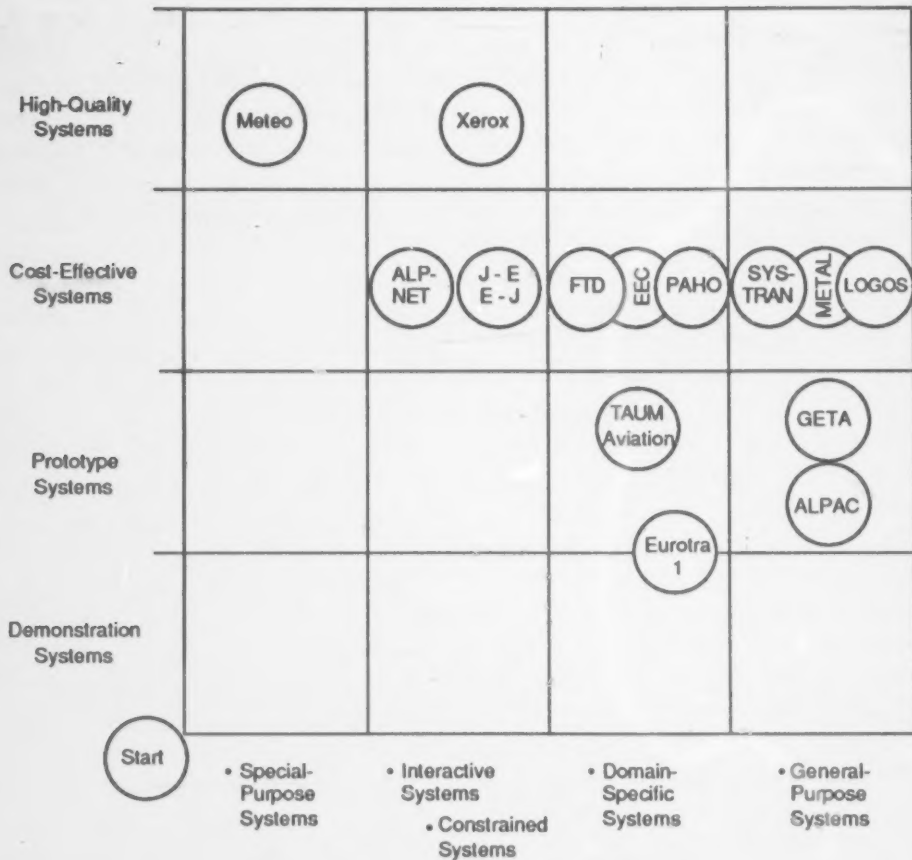


FIGURE 1 Machine translation quality/function matrix. SOURCE: Bernard Scott, Logos Corporation. See legend on page 6.

The argument may be rephrased as follows: Japanese to English machine translation systems can be used effectively for carefully targeted purposes and it will only be possible to improve these systems if we are willing to put resources into development and experimentation. When we also consider the growth of machine aids for human translators, such as dictionaries and other composition tools, it is clear that machine translation technologies have practical uses today, even if the dream of a fully automated, high-quality, low-cost, general-purpose system remains over the horizon.

Given this picture of promise and problems associated with machine translation, we need to examine the challenges that lie ahead for U.S. industry and the U.S. government. One set of challenges is commercial and relates to the fact that while a number of Japanese companies are working on Japanese to English machine translation systems, little similar work is going on in the United

LEGEND TO FIGURE 1

ALPAC	Russian-English machine translation systems being developed under U. S. Department of Defense funding prior to the publication of the ALPAC Report in 1966, which led to their discontinuation.
ALPNET	Interactive multilingual translation aid, developed in the United States. Previously known as ALPS for Automated Language Processing System.
EEC	European Economic Community.
EUROTRA I	Large-scale multilingual machine translation prototype effort sponsored by the European Community. Eurotra I is intended to lead to a full-scale industrialized version, Eurotra II, encompassing 72 language pairs.
FTD	Russian-English and German-English SYSTRAN-based machine translation systems of the Foreign Technology Division of the U.S. Air Force. FTD's installation in the late 1960's represented the first operational use of machine translation.
GETA	Machine translation system for various language pairs entailing French, developed at the University of Grenoble, France. (GETA - Groupe d'Etudes pour la Traduction Automatique.)
LOGOS	Multilingual machine translation system developed and marketed by Logos Corporation, a U.S. company. Language pairs are: German-English, -French, -Italian; English-French, -Spanish, -German, and -Italian.
METAL	German-English, English-German machine translation system originally developed at the Linguistic Research Center of the University of Texas and later under joint development with Siemens AG in Munich.
METEO	English-French machine translation system developed for the nationwide weather communications network of the Canadian Meteorological Center. METEO is a derivative of the TAUM system developed at the University of Montréal.
PAHO	English-Spanish/Spanish-English machine translation systems of the Pan American Health Organization designated as SPANAM and ENGSPAN, respectively.
SYSTRAN	Machine translation system for many language pairs, developed in the U.S. and controlled by French and Japanese interests. SYSTRAN is a much enhanced derivative of the early Georgetown Automatic Translation (GAT) system.
TAUM Aviation	An English-French adaptation of TAUM technology to a specific subject matter domain or language subset, developed at the University of Montréal. (TAUM - Traduction Automatique de l'Université de Montréal.)
J-E/E-J	Japanese-English/English-Japanese machine translation systems. There are about a dozen such systems in Japan, either on the market or being prepared for the market.
XEROX	English-multitarget machine translation system developed for XEROX Corporation by the developers of SYSTRAN, and used by XEROX to translate technical manuals written in controlled English.

States. A second set of challenges includes technical problems—issues relating to the choice of R&D focus and problems in evaluating system performance. The third set of challenges may prove to be the most pressing: defining R&D policy (either at the company or the U.S. government level). Without a clear consensus on what constitutes the “state-of-the-art,” or comprehensive data on market prospects and user needs, forging an appropriate policy response is not an easy task.

The sections that follow address each of these three sets of challenges in turn, highlighting areas of uncertainty and issues that deserve further study and debate.

2

The Commercial Challenges

CURRENT STATUS OF MACHINE TRANSLATION DEVELOPMENT

How, where, and by whom are machine translation technologies and machine aids for translation being developed? Who is using them and for what purposes? Before examining Japanese to English machine translation as it is being developed and used in the United States and Japan, it is helpful to review the broader context of machine translation development throughout the world.

Machine translation development has advanced rapidly in the past few years, and while numerous challenges remain, machine-assisted translation is no longer a dream, but is actively and increasingly used around the world. Companies and governments are developing and/or using machine translation technologies in the Americas, Europe, Japan, and in the newly industrializing countries of Asia. Machine translation and machine-aided translation are now being used in organizations as diverse as translation bureaus, multinational corporations, and government defense departments. These technologies help to scan information, collect intelligence, translate product manuals for export, and improve the efficiency of professional translators. The overview below provides a basis for understanding the organizational, political, financial, and societal reasons for differences in their development and use in the United States and Japan.

The major systems in use today are not, however, aiding translation from Japanese to English; most are performing what is considered by many the easier task of translating from one Western language to another. Many examples can be found in the public or government sector. The Pan American Health

Organization, for example, uses its internally developed system to translate health and agricultural information between English and Spanish. The U.S. government, including the Air Force's Foreign Technology Division, has long used Systran to translate Russian, German, and French into English for intelligence purposes. The government is currently supporting Systran's development of a Japanese language system. Similarly, the Canadian government uses the Logos system for translating English to French at the Departments of National Defense and State, among others, and uses METEO to translate weather forecasts. The Smart Translator is used to announce job openings. The European Community is supporting a large initiative aimed at developing a system to translate among all the official languages of the Community.

In the private sector, manufacturing companies are using machine translation technology to produce product manuals and translation companies are using it to improve the volume and efficiency of translation efforts. Xerox, for example, uses the Systran machine translation system to translate its photocopier product manuals for distribution throughout the world. A European computer maker, Nixdorf, uses the Logos system to achieve the same end. Translation companies such as Lexitech in Canada also use Logos, and ALPNET has developed a worldwide network of translation services based on its own internally developed interactive and machine-aided software. METAL has a number of clients in Europe and ATAMIRI is used by Wang for the translation of product manuals into several languages.

There are a few examples of machine translation technology used for Asian languages. Logos, one of the oldest U.S. machine translation companies, got its start in 1971 when it developed an English to Vietnamese system for the U.S. Defense Department. Currently, the Defense Department is funding machine translation from and into Korean.

It can be seen from the examples above that the world of machine translation spans the globe and includes efforts in many language pairs for a variety of uses. Some of the oldest, most successful, and well-established machine translation companies are American or are based on technology developed in the United States. Systran is perhaps the grandfather of U.S. machine translation companies and its efforts in language pairs other than those involving Japanese continue to be U.S.-based. The next oldest commercial machine translation system is that of Logos, also an American company. One of the best-known Japanese machine translation companies, Bravice International, which claims to have sold 4,500 software units, bought its technology in the United States through the acquisition of a U.S. firm. It is useful to consider this context as one examines the more specific case of Japanese to English machine translation.

Who, then, are the developers of machine or machine-aided translation between Japanese and English? The major commercial efforts in the United States are being conducted by a handful of companies. As mentioned above,

Systran, with U.S. government support, is supporting the Japanese-English combination; it should be noted, however, that Systran's commercial Japanese efforts have been Japanese-owned. Some large U.S. corporations, including IBM and DEC, are pursuing internal research in this field, but their Japanese to English efforts are located in Japan.⁷ IBM's English to Japanese machine translation system, developed for internal use in the corporation's Tokyo facilities, is currently in operational use for the translation of computer manuals.

There are some small-scale efforts, experiments in assembling and using machine translation technologies for very limited domains. Examples are Smart Communications and a small ongoing effort at the Veterans Administration Medical Center in Baltimore that uses public domain software to translate a narrow range of medical documents between Japanese and English. Efforts in machine-aided translation development are currently underway at ALPNET and LinguaTech, both of which are based in Utah and originated at Brigham Young University. Logos, which continues to invest heavily in machine translation development, has conducted research in Japanese but has never undertaken development of a Japanese system.

There is also considerable theoretical research under way among computational linguists at U.S. universities, such as work at Carnegie Mellon University's Center for Machine Translation, and the Linguistics Research Center at the University of Texas, and the sub-language approach pursued cooperatively by Hunter, Monmouth, and N.Y.U. If basic research on natural language processing and computational linguistics is taken into account, the United States still maintains a significant research effort.⁸ What is lacking in the United States is a strong development effort on Japanese to English machine translation.

The world of Japanese to English machine translation development in Japan offers a sharp contrast to that in the United States. Every major Japanese computer or electronics firm has invested considerable effort in machine translation research and development and many claim to have introduced workable systems. Without evaluating quality, it is nonetheless significant that there are at least twenty operational⁹ systems in Japan that translate from English to Japanese. Operational systems in Japan that translate from Japanese to English

⁷ At its Tokyo Research Lab, IBM is working on a Japanese to English machine translation system for use in translating newspaper editorials and economic materials, which is now in a research prototype stage.

⁸ It should be noted, however, that research in computational linguistics that is not related to machine translation will not necessarily contribute to machine translation. On the other hand, the case can be made that work on machine translation serves as a test bed and stimulus for other kinds of natural language processing investigation.

⁹ There is considerable ambiguity about what constitutes an "operational" or "usable" system. The sheer number of Japanese developers who even claim to have an operational system is nonetheless an indication of the comparatively strong interest and resources devoted to machine translation in Japan.

are available from NEC, Fujitsu, Oki Electric, Bravice International, Sharp, Toshiba, Hitachi, and Sanyo Electric to mention a few of the companies. Japanese to English systems are also under development at NTT, Mitsubishi Electric, KDD, and Toshiba.

In addition, the Japanese government has supported an important effort in machine translation development. This effort, which involves the Ministry of International Trade and Industry's (MITI) Electrotechnical Lab, the Science and Technology Agency's (STA) Japan Information Center of Science and Technology (JICST), and Kyoto University, was started in 1982. The Electronic Dictionary Research Project conducted by MITI in connection with the Fifth Generation Computer Project aims at the development of a detailed dictionary with more than 200,000 words and multiple usages.¹⁰ Supported by the Ministry of Post and Telecommunications, research is underway at ATR on speech translation telephony.¹¹ Technology transfer to industry has been made possible throughout these projects via industrial participation. The projects feed into the effort at JICST for translating science and technology abstracts. These projects are developmental vehicles that spin off nationwide results; their continuity and commercial emphasis build capability in the companies.

Who, then, is using Japanese to English machine translation? If the typical pattern is for translation to be done in the country of the target language, then Japanese to English machine translation should be done most efficiently in the United States, rather than Japan. At the present time, however, this is not the case. Most Japanese to English machine translation is conducted in Japan. The fact that many Japanese computer makers have developed their own machine translation capability reflects their orientation toward product exports and their need to control the quality of translated manuals. It also reflects an understanding of the importance of machine translation technology and its possible spin-offs to the information industry as a whole.

Japanese-developed Japanese to English systems are not widely used in the United States. Lack of compatibility between hardware and software is one of the impediments. Microelectronics and Computer Technology Corporation (MCC), a private U.S. microelectronics and computer science consortium, is a relatively new user of Japanese to English machine translation (MCC is using a Japanese-developed system) and has begun to use the technology in a particularly forward looking manner that will be discussed in more detail below. The University of Wisconsin's Biotechnology Center planned to use a Bravice system

¹⁰ The 10-year EDR project that began three years ago includes an English dictionary, a Japanese dictionary, and a neutral dictionary that connects the Japanese and English dictionaries.

¹¹ See Hitoshi Iida, "Advanced Dialogue Translation Techniques," ATR Interpreting Telephony Research Laboratories, ATR Symposium on Basic Research for Telephone Interpretation, December 11-12, 1989.

on an experimental basis to scan and translate Japanese language databases on biotechnology, but has been unable to do so because of the high cost of the necessary hardware. Currently, there is no U.S.-developed Japanese to English machine translation system on the market. The U.S. defense and intelligence community will, in all likelihood, be the first major user of Systran's Japanese to English system when it is fully operational.

MARKET PROSPECTS

Today, the volume of machine-translated documents remains comparatively small. In contrast, some Japanese believe that the annual market for all translated materials in 1988 was about 800 million yen, and that the quantity of translation will double over the 1990-1992 period.¹² One developer even estimates that by the year 2000 there will be 500,000 to 2 million machine translation systems in use throughout the world,¹³ assuming that substantial improvements are made in intermediate systems that can run on small personal computers. According to this vision, international businessmen will need small computers with built-in machine translation systems.

At the same time, Japanese experts note that large Japanese companies working on machine translation do not believe that this business will yield great profits, at least in the short term. They do, however, see machine translation as a mechanism for learning more about natural language processing technology in general, which they judge to be a key technology in the next century.

Despite the interest in machine translation technology, profits in the United States and Europe are very slim, if there are any at all. While there are companies developing systems for internal use, the independent developers are (as noted above) few in number. This contrast with the situation in Japan may be explained in a number of ways. Critics of machine translation argue that the products are based on dated technology. They argue that translators harbor serious doubts about machine translation on quality grounds, even if translators have a hard time quantifying the concept of quality. Translation, particularly in Europe, is poorly integrated with office and publishing computer environments where the potential benefits of machine translation could be substantial.

From a Japanese viewpoint, however, it is possible to create demand. Viewed from this angle, the more machine translation systems are made available and put into practical use, the more the demand will grow. Japanese industry and government are willing to plow large investments into Japanese to English

¹² See Japan Electronic Industry Association, *A Japanese View of Machine Translation*. . . , op. cit., p. 5 and Appendix 5.

¹³ Many experts see this as a very high estimate.

machine translation. It is estimated that a four- to five-year effort involving 50 to 70 people is needed to develop a general-purpose mainframe system for delivery, at a cost of \$13 million for the entire period.¹⁴ This is, however, only the beginning of the investment that is necessary. After delivery of the system, considerable resources must be invested in maintaining and improving it in response to user complaints and needs. In short, the required investments are so large that most companies find it impossible to recover costs by selling only hundreds of systems. This is a primary reason for Japanese government support not only of researchers, but also of commercial developers, although there are other important reasons such as reducing the language barrier between Japan and other countries and disseminating Japanese technical information worldwide.

Given that the major commercial machine translation systems have been mainframe-based, the traditional markets for machine translation have been limited to translation bureaus, multinational corporations, and intelligence and information gatherers, particularly in government. Developers have identified two major targets for development. One is the large-scale, general-purpose, mainframe-based system for use by big companies and governments. The goal of this type of machine translation system using large-scale hardware is high volume translation of documents, sometimes for mass distribution. Even those involved in development work on general-purpose, publication quality machine translation systems say that such high-quality, general-purpose systems are 10 to 20 years over the horizon.

The second is a small-scale machine translation system for use by small to medium size companies and even individual researchers. The purpose of this type of smaller scale system is to translate for specialized applications. Many Japanese commercial developers are focusing on small-scale system development. There is a third kind of development that is intermediate between the two systems mentioned above.¹⁵ From a commercial perspective, it is this kind of intermediate system development that is seen by the Japanese as the only feasible target for at least the next decade. Such systems are in use today for information scanning in restricted domains.

In developing and enhancing machine translation systems, developers stand to benefit from close interaction with users. Japanese developers rarely consult with users on the details of systems design, but they do seek out users' views on what features are needed in pre- and post-editing and on general issues of man-machine interface. Japanese developers prefer to interact intensively with a limited number of users so that they can respond effectively to their needs. A

¹⁴ Makoto Nagao, written response to "stimulation questions" prepared for the Symposium on Japanese to English Machine Translation.

¹⁵ Systems like those at Xerox run on microcomputers and are specialized, but cannot be considered "small-scale" since they translate more than 10 million words annually.

conscious strategy of selling a limited number of systems is often pursued in order to facilitate this process. New users benefit later from this accumulated feedback embodied in system configurations developed and perfected for other users. Although Japanese observers complain that mechanisms for exchange of information between developers and users are inadequate, in comparison to the situation in the United States there has been closer interaction among the research, development, and user communities in Japan.¹⁶

One of the limiting factors on the machine translation market is the fact that most Japanese systems operate on only one type of hardware. In fact, some large mainframe computer makers pursue machine translation development as a strategy for marketing their hardware to large companies. Bravice produces the only Japanese to English machine translation system that can be used on a variety of hardware. It runs on most small microcomputers sold in Japan. A practical barrier to widespread usage of small-scale Japanese-developed machine translation systems is the limited availability of hardware that supports Japanese characters (*kanji*, *hiragana*, *katakana*) outside of Japan.

International competition in machine translation is not mature. Given the barriers to hardware interoperability, the most prominent examples of competition occur among large-scale communication systems. In 1988, a number of computer companies competed for contracts associated with the development of a large communication system for the Korean and U.S. armies and some machine translation system developers received contracts.

Reflecting the practical limitations on machine translation technology today, there are a number of efforts that focus on combining machine translation technology and human translation. ALPNET has developed a strategy based on the assessment that machine translation is not the solution to all user needs, and that it is not an all-or-nothing alternative. Technology tools now in existence (such as multilingual word processors, dictionary look-up systems, character recognition and word processing systems) can be effectively used by human translators at much less cost than fully automated machine translation systems. New "linguistic engine" options can be added to a user's existing applications, thereby increasing ease of use without forcing the customer to learn new user interfaces to complex systems. The experiences of developers like ALPNET illustrate the fact that successful application of machine translation and related technologies depends on an understanding of user needs, establishment of expectation levels, matching the tools to the job, and the training of skilled professional people to use the tools.

Another possible approach is to redesign systems using new technology and locate smaller, modular systems in places such as schools where translators are

¹⁶ Exceptions to this general statement about interactions in the United States include relationships between developers and users such as Systran and the U.S. Air Force or between Logos and AT&T.

being trained. Also, a broader appreciation of machine translation and related technologies to cover a wide range of communications problems could enhance the use of currently existing tools in solving practical problems. For example, many potential users in Europe are not professional translators but secretaries working for large multinational corporations who must write correspondence in foreign languages. Market prospects could be broadened by integrating linguistic tools into the office environment, the publishing business, and even the engineering development environment.

If we consider particular applications of machine-aided translation technologies to specific consumer products, demand may indeed grow significantly over the next few years. A composition aid for technical writers, for example, could improve understanding of cultural nuances, not to mention grammar and syntax. Another product application that illustrates the point is a hand-held electronic phrase book for use by foreign travelers. Such applications, it should be noted, offer promise, but they remain distant from the machine translation heaven of general-purpose, high-quality systems.

USER NEEDS

People who want to read translated material care little whether the translation is done by people or machines or a combination. They simply want reliable translation that is cheap and fast. For some of these users, such as those who must scan very large volumes of information in order to follow trends in research and development or identify documents for full translation, the speed of translation is important. Turnaround time is, in the experience of the users of the Air Force's system, more important than quality, defined in terms of naturalness of expression and precision in conveying meaning.¹⁷

Other large volume users, such as companies that require translation of their operating manuals, require precise and understandable translation—but in fairly limited domains. The classic example of such a user in the United States is Xerox. Xerox has been using machine translation for more than a decade to translate technical documents. These are large volume projects which involve high reproduction and updating rates. For large companies that sell a wide range of products in global markets, translation costs represent a significant component of new product expenditures.

The Xerox experience deserves further mention. Xerox has developed an approach that integrates desktop editing programs and a Systran machine translation system to produce service manuals, training programs, and operator

¹⁷ Some argue that quality standards must include a mixture of turnaround time, accuracy, and readability, the relative importance of each varying with the needs of the particular user.

manuals for use in Europe, Latin America, and North America. This approach, which has evolved over time, involves the training of technical writers and pre-editors in a set of simple writing tools (Multinational Customized English) developed by Xerox. Dictionary building is constantly under way as technical writers and translators from different parts of the world, joined through the Xerox worldwide network, submit new words and phrases.

At Xerox, machine translation has shown concrete results. Significant improvements have been seen in productivity per finished page of translated text. Producing more than 40,000 pages annually of documentation translated through this process, Xerox now finds that translation is no longer a barrier to product launch. Machine translation enhances the company's ability to introduce new products almost simultaneously all over the world.

Xerox found that management initially had high expectations for what machine translation could do, while linguists were skeptical. Over time and with growing experience, the perceptions of these two groups have begun to converge. (See Figure 2.) By gradually developing a system that translators are able to see as a tool in their work, and one that produces high-volume product in restricted technical domains, machine translation has been fully integrated with business operations.

In contrast to large users, there are countless potential users of small machine translation systems who have different kinds of needs. In many instances, these are researchers (scientists and technical personnel) in the United States who need to know about developments in Japanese science and technology. For many of them, readability and smoothness of translation is less important than timeliness and access to a narrowly targeted range of technical documents of interest.

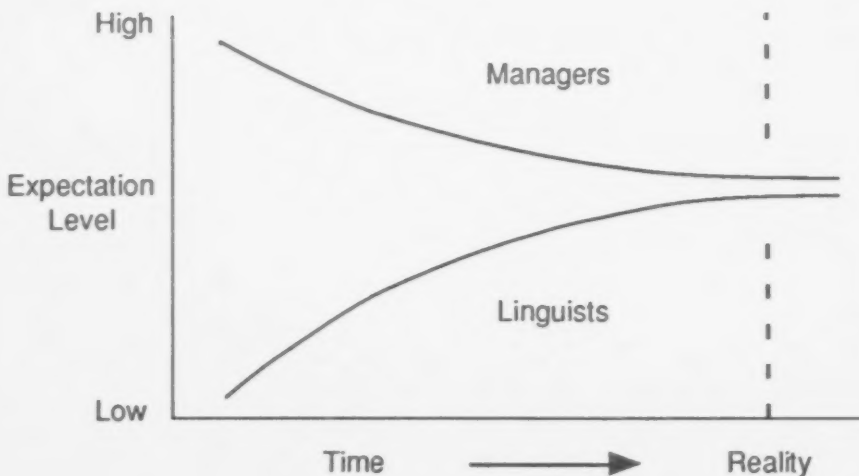


FIGURE 2 Expectations versus reality of computer translation. SOURCE: Maria Russo.

MCC is now experimenting with machine translation to monitor overseas technology developments and provide translated materials to member firms. For an investment of \$50,000, MCC has assembled a machine translation system that includes an optical character reader, a Japanese-made personal computer that is also used as a JICST terminal, a Japanese to English machine translation software package, and a U.S.-developed workstation. MCC encountered problems in connecting the computer systems developed in Japan and the United States and in getting adequate vendor support for the machine translation system developed in Japan. Fifty percent of the sentences in the trial output have been judged to be accurate. For a modest investment (\$50,000 to \$60,000), the machine translation experiment at this U.S. industry consortium is considered to be worth the effort.¹⁸ MCC's International Liaison Office expects to augment its already successfully established functions of monitoring developments in Japan and providing technical support to its researchers by using machine translation first to translate titles and, eventually, abstracts and short papers. MCC has, in addition, recently initiated a research effort on knowledge-based natural language processing targeted toward machine translation technology development.

Beyond corporations, individual professional translators and researchers also qualify as potential users of machine translation systems and machine aids for translators. Companies such as LinguaTech are developing workstations for translators who can work at home using a microcomputer with a high-capacity disk drive, a printer, and a port for telecommunications. A translator can receive source texts and glossaries with specific subject matter along telephone lines. A variety of options are available to the translator, including composition tools, a bilingual corpus that permits text retrievals when necessary, multilevel terminology files, and optional use of machine translated texts. To use a baking analogy, machine translation is one ingredient in a range of elements that the translator (baker) can combine in myriad ways to accomplish his work. The baker may wish to use the heavy-duty bread mixer (machine translated text) for some tasks, but he also keeps his drawer of spoons (manual tools).¹⁹

Potential users, particularly individuals or small companies, may find it hard to learn about machine translation. In Japan, potential users learn about machine translation through newspapers and television as well as through demonstrations at computer company service centers. Only limited information is obtained through these channels. The potential Japanese user is, however, in a much better

¹⁸ In judging the value of machine translation, it must be remembered that, unlike the case with European languages with cognates and familiar script, Japanese is totally unintelligible to the potential English-speaking user. Some, therefore, believe that capturing even 50% of the Japanese text can be a useful step forward in communication.

¹⁹ This analogy is developed by Alan K. Melby in his article "Recipe for a Translator Work Station," *Multilingua*, 3-4 (1984), pp. 225-228.

position than his U.S. counterpart. For the U.S. translation community, demonstrations and meetings of professional organizations provide channels for transmitting information that augments the articles appearing in specialist journals. For other users who have no direct contact with the translation community (such as researchers and people in small business), acquiring information about machine translation and making judgments about purchase and use are even more difficult. We shall return to this point again in the final section on R&D policy.

JAPANESE AND U.S. USERS: CONTRASTING NEEDS

As we think about what type of Japanese to English machine translation system would help users in the United States, it is important to keep in mind that their needs are different from those of most users in Japan. The profile of the typical user of Japanese to English machine translation in Japan is a large corporation engaged in global exports. This user can "control" the source document because it is in his native language and pre-editing is easier in this case. This user can customize the system to suit his or her needs and tolerate marginal machine translation because there is a direct cost justification for this effort.

The profile of the typical user of Japanese to English machine translation in the United States is someone who needs expanded access to Japanese-language technical information but who has no fluency in Japanese. This user has no control over the source text and little ability to pre-edit. For monolingual users, the "raw" output must be more reliable and accurate than for the typical Japanese user. For the user in the United States, there are many problems associated with gaining knowledge of and access to Japanese databases, inputting text, and post-editing requirements. In view of the broad-based needs of the U.S. user, marginal machine translation is in many cases not useful. (See Figure 3.) However, there are many possible uses of machine translated text now available, particularly if some post-editing is done, for scanning information.

Japanese to English machine translation systems developed in Japan do not, in the opinion of some leading U.S. developers, fit the needs of many potential high-volume U.S. users. The requirements are for high-quality, broad-domain systems which will be based on new technology.

What are our choices? Should we wait for Japan to develop high-quality Japanese to English machine translation systems in the year 2000 and later? After all, Japan has a head start and Japanese is their language not ours. Or should we develop Japanese to English machine translation systems here in the United States? Are there other alternatives? Answering these questions requires an understanding of the technical challenges facing those engaged in R&D on machine translation.

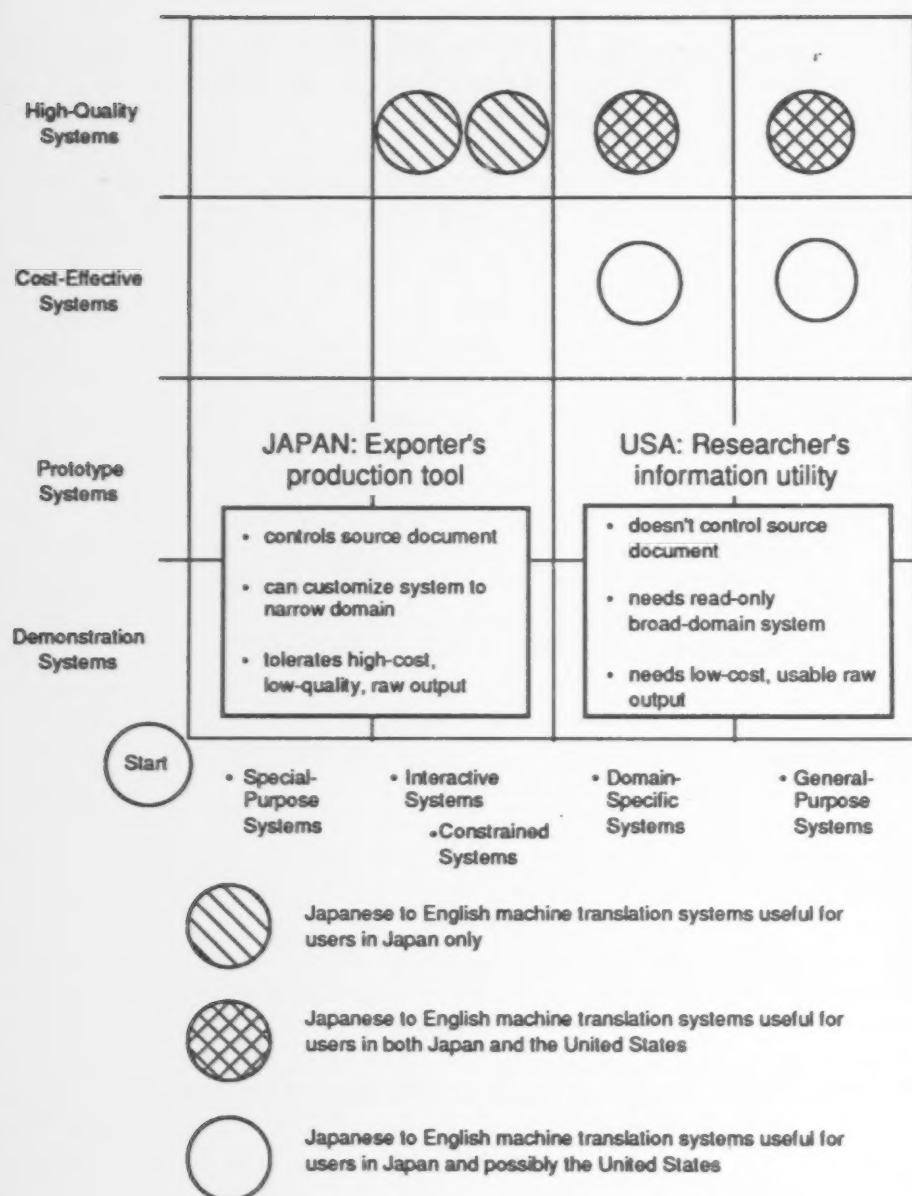


FIGURE 3 User profiles for Japanese-English machine translation. SOURCE: Bernard Scott, Logos Corporation.

3

The Technical Challenges: Approaches to Research and Assessment

Dramatic improvements in computer hardware and software have contributed to progress in machine translation. System performance, which can be measured in "raw LIPS" or logical inferences per second, is now doubling every one to two years.²⁰ Despite, or perhaps because of, these rapid advances in computer technology the barriers in linguistic theory and other areas have become ever more apparent. The result is a receding horizon—as strides are made in R&D it is clear that much more needs to be done.

Over more than 30 years of research and development, work on machine translation has taken three approaches to the process of translation. But there are a variety of machine translation systems in use today and new advances in technology have ushered in systems using intermediate language representations and artificial intelligence that enable the computer to "learn" a language. The discussion that follows briefly reviews the current strategies for developing machine translation systems, key problems for research and development, and issues in the evaluation of machine translation systems.

²⁰ Logical inferences per second is a measure of the speed at which the system: 1) recognizes a pattern match between an element of the input text and a previously stored pattern, and 2) applies the rule that goes with that pattern to translate the element.

DEVELOPMENTAL STRATEGIES AND PROBLEMS

Regardless of which languages are translated, there are now three primary translation strategies for machine translation.²¹ The direct translation method deals only with single language pairs and translates words directly from one language to another. Used for most of the earliest systems, this method involves very little or no linguistic analysis and produces very rough translations. Although this strategy is not designed to handle translations of complete documents, it has been used for machine translation of large databases, tables of contents, and titles of technical publications. The International Liaison Office of MCC is using this strategy for Japanese to English translation in order to develop databases on scientific developments in Japan for its customers who require an overview of available materials in particular fields. It is expected that the users will select documents for full translation by other means.

The transfer system, which operates in three stages, is the most widely used strategy for machine translation. The source language is first analyzed and converted into representations that can be transposed into sentence structures through semantic analysis. In the second stage, the source language representations are converted to the target language transfer representations. The final process synthesizes the transfer language representation into the text of the target language. This method is the one most widely used in Japanese to English machine translation systems, although the transfer system works best when the language pairs are closely related. In some mainframe systems such as experimental systems ATHENE/N developed by Hitachi and FAI by Mitsubishi, the transfer system is enhanced by the use of case analysis.²²

The pivot method is the third translation strategy. It is based on the ideas that language is a universal human experience and that a universal interlingua can be developed, which can be understood by a machine. This method is designed to convert the source language into the interlingua which is then converted into the target language. The interlingua today is still largely a theoretical concept, although the Logos system makes use of an interlingua in a hybrid interlingual/transfer architecture. Researchers working on the interlingua expect that the application of artificial intelligence will permit significant advances to be made.²³

²¹ This overview of machine translation strategies is drawn from a paper by Wayne Kiyosaki, "Machine Translation: Time for a Reappraisal," forthcoming.

²² "Japanese Machine Translation Systems Described," Tokyo NIKKEI ELECTRONICS in Japanese, February 1986, pp.137-168.

²³ For a detailed analysis of the strengths and weaknesses of these three strategies, see W. John Hutchins, "Recent Developments in Machine Translation," *New Directions in Machine Translation*, Conference Proceedings, Budapest, August 1988.

The systems that have been developed use various approaches; they can be compared to tools in a tool chest. No one tool is always best but in some cases one tool may be better than another.²⁴

Using primarily direct translation or translation by the transfer system, there are three possible ways in which human intervention can occur. Pre-editing can involve two kinds of operations. In one case, a text is revised to eliminate structural or lexical ambiguities before being translated by a computer. In the past this approach was not widely used, due to the difficulty in anticipating structures or words that will be difficult for a computer to handle. More recently, with the introduction of text-critiquing software, the potential ambiguities can be brought to the attention of a human translator automatically.

In a second approach to pre-editing, the input text is produced especially for the machine. In some cases it is a new version of an existing text, in others an entirely new text. Multinational Customized English is an example of a restricted English developed by Xerox for use on its Systran system. In some cases, pre-editing is almost as difficult as traditional translation. The efficacy of pre-editing depends to a great extent on the human editor knowing the limitations of the machine translation system.

In the case of interactive editing, the computer calls on the editor to make choices among various alternatives in order to resolve ambiguities. It is also possible to combine post-editing with interactive editing. However, this can make the process costly. The first interactive systems were introduced in the early 1980s and the interactive editing method seems to have gained wider acceptance in recent years.

Of the three editing options, post-editing is clearly the most widely used. Usually a professional translator, the post-editor corrects the machine's output. This is more efficient when done directly on the screen using appropriate word processing software. If the post-editor writes the corrections on hard copy and they are then entered into the computer, the process is much slower. Some estimate that an experienced post-editor can produce 4,000 to 8,000 words a day and in some cases as many as 10,000.²⁵

Since human intervention is costly, the goal of some developmental efforts is fully automatic operation of a machine translation system.²⁶ When the application involves merely gleaning the "gist" of the text, some of the large, general-purpose systems are used on a fully automatic basis. If a more careful translation is needed, output can be post-edited. Such systems include general-

²⁴ Observation by Jaime Carbonell, Carnegie Mellon University.

²⁵ This discussion of editing options is based on work by Muriel Vasconcellos.

²⁶ The Center for Machine Translation at Carnegie Mellon University is working to improve the quality of machine translation output through the incorporation of knowledge bases, especially for applications in limited domains.

purpose systems that are able to handle a wide variety of source texts and special-purpose systems designed to translate a special type of source text such as weather reports or abstracts of technical articles in particular fields.

In addition to full machine translation systems, there are many related technologies that are used as translation aids. These include on-line dictionaries, grammar checkers, and libraries of phrases that are regularly used by human translators. These systems are updated and developed as post-editors contribute to the on-line dictionaries and users give input to improve the way that the machine translation system does the actual translation.

Research and technical challenges particularly relevant to Japanese to English machine translation include the problem of inputting the text, which includes Chinese characters as well as two phonetic scripts. Optical character readers will help to solve the problem of text input, but there are still many difficulties associated with input of Japanese text because of different character fonts and the placement of charts, graphs, and tables in the text. Optical character readers are now being coupled with machine translation systems in Japan, but the extent to which they increase savings over manual input is not clear.

A major research and development question relates to the problem of pre-editing. The better the source text (the clearer the expression and the shorter the sentences), the better the resulting machine translated text and the less post-editing needed. But, as mentioned above, pre-editing is time consuming and tedious work that requires special skills.

While significant advances have been made in computational linguistics, there remain problems that must be overcome in order to build linguistic theory and develop more sophisticated machine translation systems. This set of challenges could be approached in a step-by-step fashion, as some Japanese experts suggest. Research in the following areas is needed: the introduction of priority information in order to disambiguate several possible sentence structures and words; the development of learning mechanisms that produce preference values for the disambiguations; the establishment of grammatical rules that consider many more than two elements simultaneously; improved capabilities for dealing with such problems as anaphora resolution, ellipsis, and the analysis of sentence fragments.

Although machine translation strategies and system types are more or less universal, the ways in which the researchers in the United States and Japan approach these subjects are quite different. As one observer put it: Americans write papers; Japanese build machines.²⁷ The Japanese approach has been more

²⁷ This distinction should be considered carefully. Some question the notion that Japanese researchers are not theory-oriented: one leading Japanese researcher believes that (instead) they focus on second and third approximations required for machine processing of natural language—less beautiful and less academic but useful theory. On the other hand, one critic of Japan's machine translation says that the machines that the Japanese build do not really do the job and (therefore) their approach is not practical.

pragmatic and oriented toward experimenting with systems. This involves a problem-solving approach to linguistic analysis. The parts of the language that do not fit neatly into linguistic theory models are approached by combining different theories or by accumulating individual facts to deal with specific problem areas.

In contrast, the U.S. research community has concentrated more on machine translation theory than on applications. Much of the researcher's time is devoted to writing papers and developing new models of natural language. As a result, critics argue that U.S. researchers construct models that are elegant but not amenable to practical use. At the same time, we should remember that the strides that have been made in basic computational linguistics, a research approach recommended by the ALPAC report, make today's machine translation systems possible.

The theoretical work that has been done in the United States and other countries, including Japan, has made machine translation developers and users aware of the research challenges that are present. These include the need for a bilingual text corpus and the development of automatic comparison algorithms for this corpus. The automatic collection of special terminology words and construction of a thesaurus of these terms would improve many machine translation systems. Standardizing dictionary theory and practice, proper analysis of broken utterances, improved grammar checking devices, and automated approaches to the detection and resolution of ambiguities are other important research themes. Even linguistic and cognitive studies of pre- and post-editors' behavior have been suggested as avenues to improved machine translation.

All of this requires an increase in the number of researchers working on machine translation as well as more basic research themes. Some practical steps might be taken to make experimental tools for natural language processing and machine translation easily available to researchers. These might include the construction of a portable software package for natural language processing, and its distribution to interested researchers; establishment of core grammars for English and Japanese that are linguistically sound, and their distribution to interested researchers; the construction of a text database that includes bilingual text data for use in natural language processing.

In the United States, where the thrust of research has been in more theoretical areas, there is a need to improve interactions with those who take an "engineering" and applications-oriented approach if commercialization of useful systems is the objective. As noted above, interaction with users is essential to system development. These and other questions central to R&D policy in the United States will be explored more fully in Section 4.

EVALUATING MACHINE TRANSLATION SYSTEMS

Corporations involved in development, researchers working on fundamental technologies, potential users, and government policymakers all need to know how good machine translation systems are in order to make choices. Unfortunately, there is no generally accepted method for evaluating the quality and accuracy of translations by people, or by machines.

Japanese developers of machine translation systems often say that the systems are 80% acceptable. This general score is, however, more an intuitive judgment than the result of systematic research. It was pointed out that if 20% of the cookies in the cookie jar are poisoned, no one will want to eat any of them.

Overall assessments of machine translation are less useful than evaluations of specific systems because the evaluation depends very much on the needs of a specific user. Japanese developers note that in some cases a reasonably accurate or even a rough translation may be appropriate, while in other cases where high levels of accuracy are essential, machine translation is unacceptable. A researcher who needs to comb through a vast mountain of information may find rough translations of abstracts very useful in tracking overall trends in research or in selecting articles for full translation. Nothing less than absolute accuracy in translation will satisfy a lawyer working on a legal brief or a politician whose words are quoted by the media. The machine translation systems now in operation, particularly the prototype Japanese to English systems, have been developed to translate technical documents, manuals, and information in restricted domains.

Participants in the Japanese machine translation project supported by the Science and Technology Agency of Japan developed an approach to evaluation using two independent indicators: intelligibility (the extent to which the translation can be understood by a native speaker of the target language) and accuracy (the degree to which the translated text conveys the meaning of the original).²⁸ Samples of machine translated sentences were evaluated by the researchers as roughly 80% acceptable. This overall evaluation was based on the result that 80% of the sentences were given a score of at least 3 in intelligibility and accuracy.²⁹ It is estimated that 20 to 30% of the output sentences in Japanese to English machine translation systems are unacceptable, and in those cases post-editing cannot be carried out effectively.

²⁸ See Makoto Nagao, Junichi Tsuji, and Junichi Nakamura, "Machine Translation from Japanese into English," *Proceedings of the IEEE*, vol. 74, July 1986.

²⁹ A score of 3 in intelligibility was given to sentences whose meaning was clear, but where the evaluator was not sure of some word and grammar usage. A score of 3 in accuracy was given to sentences where the content of the input sentence was generally conveyed in the output sentence, but where there were problems with tense, voice, etc. *Ibid.*, p. 1006.

While no commercially available system can do it, some Japanese to English systems now in use by researchers in Japan reportedly can identify inaccurate text. Leaders in Japanese to English machine translation research, however, note that no accurate data are available to judge particular systems and that the assessments of accuracy and intelligibility are not based on rigorous testing.³⁰

Nor are there unambiguous cost evaluations of machine translation systems, although developers contend that the time taken and cost are generally less than with pure human translation. Here, again, the conclusions drawn about the relative cost of machine translation depend on the type of text and the purpose of the user. According to Japanese expert reports, the best Japanese machine translation systems are cost effective. In one example, a page of text can be translated in 40 minutes when post-editing is done on hard copy, while human translation requires about 43 minutes per page. The charge for machine translation is about 75% the amount for human translation in this particular instance.³¹

The more a user uses a machine translation system, the more efficient the work. It takes at least one year and usually two years for a user to become really familiar with a system and for cost efficiencies to become apparent. (See Figures 4 and 5.) It appears that a significant volume of text must be translated in order to achieve such "learning curve" benefits. The more carefully selected the text (with short sentences and well tuned content consistent with the parameters of the system), the more apparent the cost efficiencies over time. (See Figures 6 and 7.)

Unfortunately, evaluations of machine translation systems currently depend on subjective judgments as to what constitutes acceptable levels of cost and accuracy. In many respects, beauty is in the eye of the beholder. What may be unacceptable text to one user may be usable to another. A major obstacle to the development of machine translation systems is the reluctance of some involved in development to provide detailed information about performance characteristics and to exchange information about their experiences. Developers anxious to convince potential funders of research and users of the systems have oversold their systems, resulting in frustration. Potential users are well advised to conduct systematic comparisons of system performance on sample texts of their own selection that are typical of the application envisaged. In order to facilitate research and development, it will be

³⁰ It should be noted that the ratings are carried out by the developers and reflect evaluations of carefully "tuned" texts appropriate to the system.

³¹ See Japan Electronic Industry Development Association, *A Japanese View of Machine Translation*. . . , op. cit., p. 12. This utilization example involves machine translation of a technical text "tuned" to the system. See also Appendix 9 of the report, *Examples of Machine Translation Use in Japan*. One participant in the symposium reports that better results for machine translation as compared to human translation from Japanese to English were recently reported at a conference in Munich.

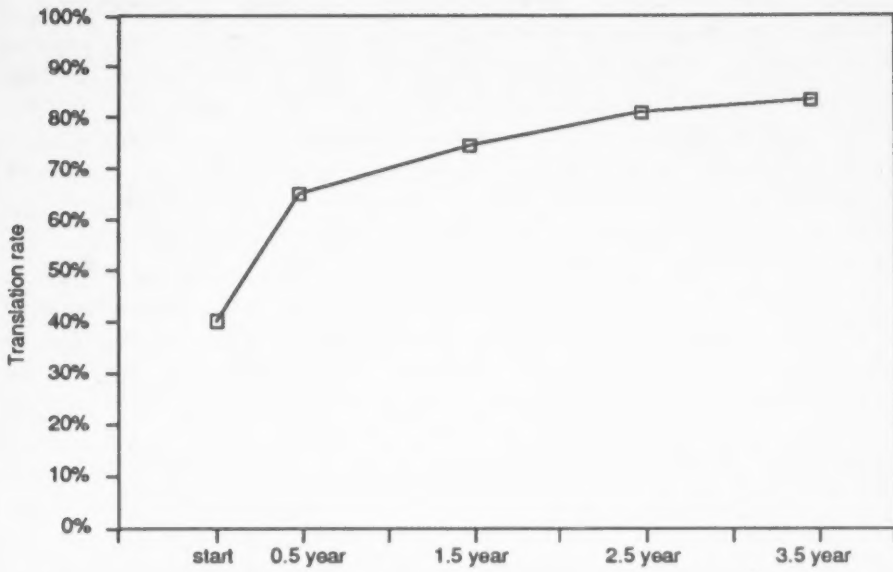


FIGURE 4 Developer's effort to improve. SOURCE: Data collected by a major Japanese firm involved in machine translation development.

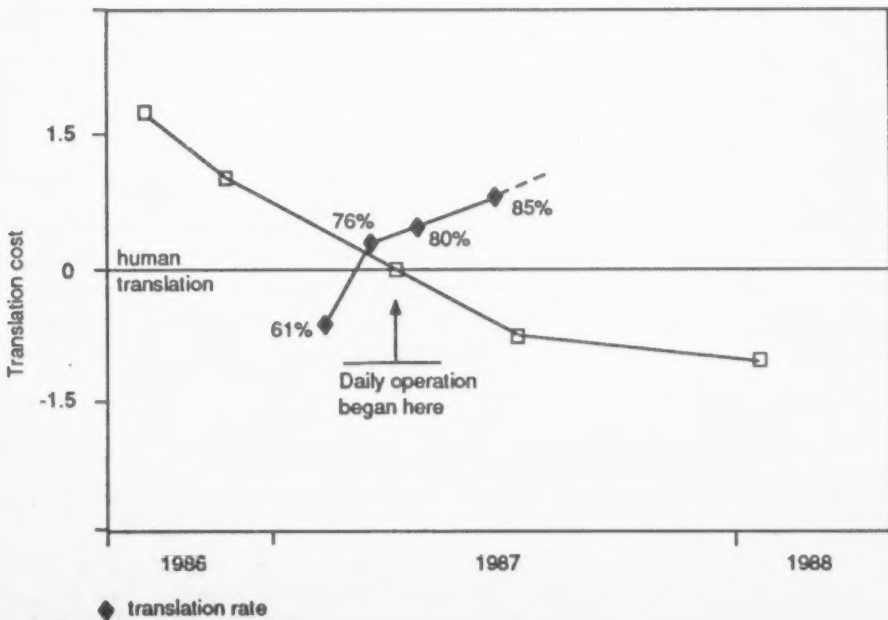


FIGURE 5 User's effort to improve. SOURCE: Data collected by a major Japanese firm involved in machine translation development.

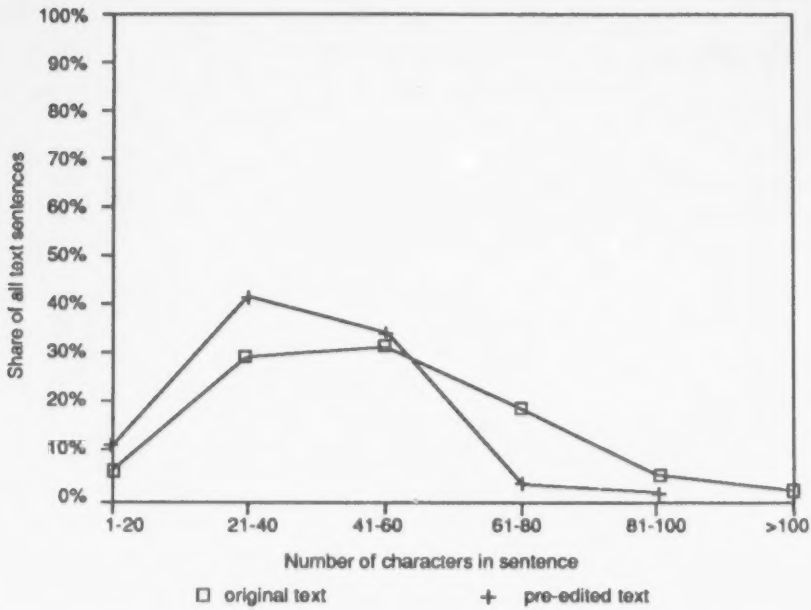


FIGURE 6 Length of sentences in text. SOURCE: Data collected by a major Japanese firm involved in machine translation development.

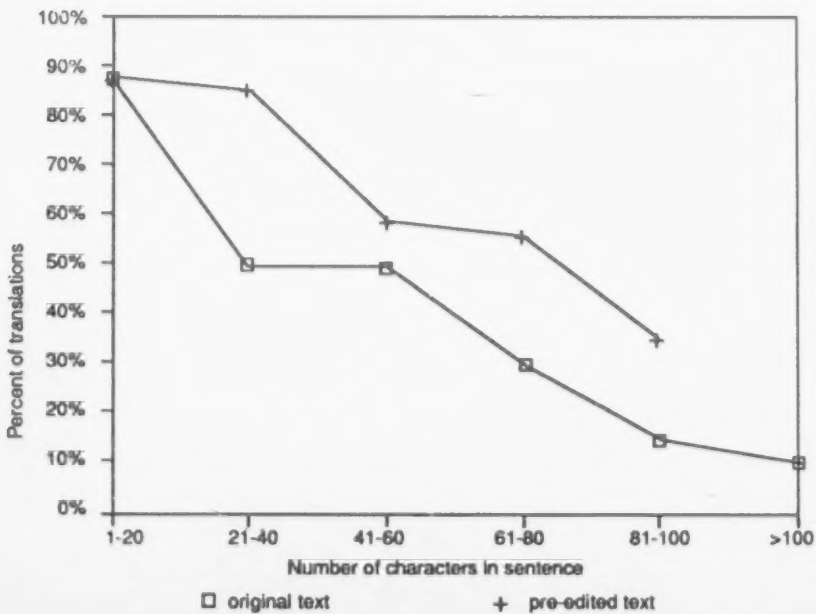


FIGURE 7 Accuracy of translation (by length of sentence). SOURCE: Data collected by a major Japanese firm involved in machine translation development.

necessary to improve techniques for evaluating system performance and timely exchange of information about new developments.³²

³² One participant in the symposium expressed doubt, based on experience of the past 20 years, that reliable methodologies for evaluating machine translation systems can be developed. A comparison of parsers under controlled conditions was suggested as a possibility.

4

The Research and Development Policy Agenda

The discussion above highlighted the special needs of U.S. users of Japanese to English machine translation for high-quality, general-purpose systems and the research and technical challenges in developing such systems. What should be done and who should do it? The sections that follow review a range of perspectives on these issues discussed at the symposium and lay out alternative approaches to the formulation of a policy agenda.

THE SKEPTICS

Despite the apparent growing need for translations of Japanese technical documents, and significant efforts by members of Congress and others to expand U.S. government efforts in this area, skeptics question whether there is real demand for such information. A number of commercial ventures to provide translated Japanese technical literature have failed and many question whether the U.S. research and business communities will overcome a "not invented here" syndrome—the assumption that technical information from abroad is second rate. At the same time many suspect that the information being made available is out-of-date and of little technical interest.

A number of observers question whether the United States government should attempt to mount a major initiative on machine translation. The case against doing something, particularly at the U.S. government level, is based on uncertainty about the need to target this set of technologies for special attention.

Despite a decline in U.S. funding for machine translation, the United States maintains a solid research base in machine translation. Skeptics, moreover, question whether Japan has really taken the lead. Instead, they say that the Japanese have rushed prematurely to market machine translation systems that need much more work and would not be attractive to the typical U.S. user. Therefore, the argument goes, there is no justification for a U.S. government effort. If the demand is there and the technical problems are amenable to resolution, they say, U.S. industry should take the lead rather than calling on the government to support costly and uncertain research.³³

Furthermore, support for R&D on machine translation must be seen in broader context. Some believe other technologies deserve more attention and may have quicker commercial payoffs. In an era of tight federal budgets, how can the case be made that machine translation deserves special attention?

Nor is machine translation a panacea for widening U.S. access to Japanese scientific and technological information. It is one piece of a large picture. Efforts to expand Japanese language study, harmonize patent systems, and reward young scientists and engineers who spend time in Japanese laboratories are also important, sometimes competing themes in the national debate. Professional translators, moreover, worry that advocates of machine translation believe that it will be possible to eliminate "the most expensive link in the communication chain"—the skilled bilingual.³⁴

Even those most committed to work on machine translation note the sociological problems that plague this enterprise. Viewed from this angle, no amount of R&D funding will by itself resolve the fragmentation of the research, development, and user communities in the United States. If machine translation is going to be supported, the skeptics say, it should be as part of a larger national commitment to building networks of interaction among researchers, business people, and professional translators.

THE ADVOCATES

In response to those who question whether the demand is "real" for translations of Japanese technical literature, advocates of new policy initiatives note the lack of follow-through in previous initiatives. Some believe that the

³³ Skeptics note that most machine translation systems in operation today are based on logic programming, the expert system approach. Marshall Unger argues that recent breakthroughs in artificial intelligence (those obtained in the neural-net paradigm, also called connectionism) offer the most promising direction for future research and development.

³⁴ As discussed above, it would appear that such concerns are no longer valid in view of changing perspectives on machine translation.

Office of Japanese Technical Literature, created by law in 1986, has been downgraded within the Department of Commerce and given inadequate funding.³⁵ Federal agencies could, using machine translation, do a better job disseminating existing data on Japanese science and technology to the public, research, and business communities alike. Machine translation, used as a gisting tool, does work that is simply beyond the capacity of human translation. If machine translation were developed into a systemized service, the result could be to stimulate new demand for Japanese technical information. If users do not know it exists or that it can be obtained at reasonable or no cost, demand will remain low.

It is not enough to say: let the market be the judge. Markets can fail to meet the national interest, providing a legitimate rationale for government action. In Japan, government and business are cooperating to support a long-term machine translation development effort. They are, in effect, creating a market. Because the needs of users in the United States are different from those in Japan and cannot be fully met by the Japanese to English machine translation systems now in operation, some believe that our needs cannot be met by simply buying from Japan.³⁶

There are broader motivations for advancing support for machine translation in the United States. In Japan and Europe, machine translation is seen as an important component in the new range of information technologies for the 21st century. As a technical challenge, Japanese to English machine translation represents an engineering stretch that may spin off progress in other areas of information technology, one that requires integration with other fields of information technology R&D. Viewed from this perspective, machine translation is a test bed for solving natural language processing problems in information technology. The question, advocates say, is whether the United States can afford not to pursue this technology.

The potential impacts of a focused effort on machine translation extend beyond information technologies. Bringing together researchers, developers, and users in a long-term project would provide an opportunity for experimenting with new approaches to R&D organization. There is a striking fragmentation of machine translation efforts in the United States—ineffective communication between scientists working on basic research and the engineers who apply it; weak links between the developers and the potential users. This lack of coordination makes machine translation a very fragile technology in the United

³⁵ Comments by Congressman George E. Brown Jr., on the "Federal Role in Accessing Japanese Technical Information" at the Symposium on Japanese to English Machine Translation.

³⁶ At least one participant in the symposium, however, argues the case for simply buying from Japan. As long as the technology is readily available to U.S. customers, the argument goes, it makes sense to take full advantage.

States today. There are also deeper issues at stake such as whether we should be satisfied with a situation in which the United States produces the best research papers and the Japanese produce the products.

WHAT MIGHT BE DONE?

In developing a policy agenda for machine translation in the United States, there are a number of levels at which initiatives can be taken. Initiatives could be taken at the international level. Second, a national policy effort could be defined. And third, individuals and organizations could define new approaches. The suggestions outlined below are not mutually exclusive, however, policymakers who see the need for action will also need to determine where scarce resources can best be invested.

International cooperation in machine translation is desirable in order to develop expanded and up-to-date bilingual dictionaries. It also makes sense in view of the large expenditures required and the "precompetitive" nature of some of the more basic research needed to lay a solid foundation for product development. Dr. Makoto Nagao, a leader of Japan's machine translation efforts, has initiated an International Association for Machine Translation that will bring together researchers, developers, and users in North America, the European Community, and Japan. The purpose of the organization would be to collect and compile information on machine translation so that users have a better understanding of what types of systems are available, how they can be used, what the experience of others has been, and improve evaluation of machine translation systems. For the developers, the benefits would include cooperative approaches to dictionary building, database construction, and sharing of texts as well as expected improvements in theories of machine translation and in machine translation systems. Users and developers alike could benefit from the development of standards for input, design of controlled languages, and evaluation. Conceived as a federation of regionally-based organizations, the international organization will be supported by membership dues (including both individual and corporate members) and other sources, including government funding.

In the context of U.S.-Japan relations, cooperation in machine translation has special significance. Machine translation has been identified by both countries as an area for potential cooperation under the United States-Japan Agreement for Cooperation in Science and Technology. Machine translation appears to be well suited for U.S.-Japan cooperation for a number of reasons. First, the precompetitive aspects of the R&D challenge are significant. Second, a long-term effort is needed. Third, the costs are high and the commercial payoffs over the horizon. Finally, the United States could potentially learn from the efforts already under way in Japan, provided that the participants include senior

researchers from Japan's premier corporate laboratories. The Japanese would, of course, benefit from opportunities to interact with the American user community and from collaboration in development of translation tools. Assuming that a U.S.-Japan cooperative project would be thoughtfully focused to ensure that the potential benefits to the United States are maximized, the significance could be broader in demonstrating the value of collaborative R&D efforts more generally.

A compelling case can also be made for a defensive United States strategy where the U.S. government takes the initiative. If consensus is built on the need for a policy initiative, the U.S. government would have to be involved. The small, struggling companies that specialize in machine translation in the United States lack the resources and the motivation to mount a serious effort in Japanese to English machine translation. For the large companies with big R&D budgets, on the other hand, the high risks and uncertain returns of expenditures on machine translation research present significant disincentives. Universities active in this field do not see it as their mission to build commercially viable, general-purpose machine translation systems. (See Figure 8.)

This leaves the U.S. government as the only potential catalyst for a serious, product-oriented effort on Japanese to English machine translation. An important requirement for such an initiative would be the participation of government, industry, and university researchers and strategists. A three-stage approach could be considered. In stage one, a steering committee would be assembled to take an inventory of technology sources and user needs. The major tasks of the steering committee would be to: determine what are the bottlenecks to the development of high-quality machine translation systems; explore the issues of evaluation criteria; develop plans for design of a new type of R&D organization. Stage one could be completed within an 18-month time frame. Stage two of a U.S. initiative, lasting 2-5 years, would focus R&D funding on the development of two prototype systems (Japanese to English; English to Japanese). Stage two efforts would be highly focused on problem solving and the evaluation of technical results. The R&D would be carried out by two competing consortia involving participation by industry, government, and university researchers. If two competing consortia were established, however, the costs would double. Another possibility would be to have two competing consortia, both working on a Japanese to English prototype system. Product-oriented implementation would be the goal of stage three, during the 5-10 year time horizon. During this stage, interim products would be developed by a number of subcontractors who would work on machine translation systems for use in a wide range of domains.

Critical to the success of such an initiative is a recognition that some of the most significant barriers to previous machine translation projects have been organizational. Involving the right mix of researchers who combine strengths in theory as well as engineering problem-solving would be essential. Building mechanisms that integrate developers and potential users at an early stage would also be important. The infrastructure already established by the National

	US Univer- sities	US Machine Translation Firms	US Industry	US Gov't	Japanese Industry *
• motivation				✓	✓
• resources					
- financial			✓	✓	✓
- organizational			✓	✓	✓
• expertise					
- core machine translation technology	✓	✓			✓
- machine translation production development		✓			✓
• user-friendly interfaces					
• user feedback					
• develop pitfalls					
• large-scale databases					
• format handling					
- integration	✓		✓		✓
• platforms					
• networks					
• database management systems					
• graphics					
• scanners					
- high-quality MT technology		?			?

* Note contributions are also made by Japanese universities and government but Japanese industry has all of these resources.

FIGURE 8 Sources for J-E development. SOURCE: Bernard Scott.

Technical Information Service for accessing and translating Japanese language documents could serve as a foundation for this effort. Machine translation could be a test bed for a focused, U.S. national initiative in information technology development. Developing procedures for rapid processing of requests to use copyrighted materials is also essential, as well as agreements that encourage timely dissemination and low cost or free usage of jointly developed dictionaries and other tools.

At another level, policy initiatives can also be developed at the organizational or corporate level. Those with long experience in providing Japanese language technical information argue that a commitment is necessary to create and maintain demand at the organizational level. MCC's International Liaison Office,

the information specialists at Bell Laboratories—these are the people and places where demand can be developed internally. Building better ways to organize, disseminate, and utilize Japanese technical literature provides, therefore, the context for machine translation. Building a constituency of users for translated documents means working directly with the users to tailor the search, selection, and output to their needs. Unfortunately, only the largest organizations today have invested the resources needed and only a few have made expanding access to foreign technical literature a high priority.

CONCLUSION

A range of options is available to policymakers (public and private) who determine that initiatives in machine translation should be pursued. These options—international collaboration with the European Community and Japan, a U.S. national project, government support for stronger linkages between university research centers of excellence and potential developers, a focused government procurement strategy designed to build machine translation expertise in the United States—are not mutually exclusive. It is, however, not clear that pursuing an international cooperation strategy would by itself build a fully integrated research and production base in the United States or a new generation of machine translation systems that fit the special needs of the U.S. user community.

Machine translation, broadly defined, offers interesting research and development challenges in building new information technologies. It also offers significant potential rewards to businesses operating in global markets, to researchers and engineers who need to know what is going on in Japan, as well as to translators whose work can be augmented by the use of machine translation tools. Machine translation is not a panacea or a fully developed technology—but it is a dynamic and challenging area worthy of serious policy consideration.

**U.S. - Japan Task Force
on
Scientific and Technical Information**

April 1990

Executive Summary

This is a report of progress made at the April 2-4, 1990 joint meeting of the U.S.-Japan Task Force on Scientific and Technical Information. The meetings were held at the National Technical Information Service in Springfield, Virginia, and were guided in their agendas by objectives set in Tokyo in April, 1989.

The Task Force is pursuing four goals:

1. Improve the awareness and understanding of organizations and systems established to improve the use of scientific and technical information. (Mutual Education)
2. Increase the quantity and quality of scientific and technical information (STI).
3. Reduce impediments, if any, to the flow of scientific and technical information. (JRRC)
4. Increase the translation of scientific and technical information. (Translation)

1. Mutual Education

Three activities are in progress: 1) Reports on the use and distribution of Japanese STI; 2) collection of videotapes describing Japanese research facilities; and 3) a conference for American researchers on access to Japanese STI.

2. Scientific and Technical Information (STI)

The Task Force mutually agreed that the term 'Grey Literature' be eliminated and emphasis be placed on exchange of S&T information. Important examples of scientific and technical information for exchange are listed in Appendix B.

Numerous significant accomplishments and plans involving JICST, the National Diet Library and NACSIS are detailed in the report.

Executive Summary

3. Japan Reprographic Right Center (JRRC)

A group of promoters composed of authors' associations, publishers' associations, academic societies and expert individuals is preparing for the establishment of the JRRC, with advice from the Agency for Cultural Affairs.

The Japanese government continues to encourage the establishment of a private Japan Reprographic Right Center (JRRC) similar to the U.S. Copyright Clearance Center.

The U.S. side strongly urged early establishment of the JRRC.

4. Translation

There is no mechanism to facilitate negotiation of translation rights between American researchers and Japanese authors and publishers. The Task Force recommends a joint study of some practicable mechanism to facilitate communication regarding STI translation rights between concerned parties.

Until recently the U.S. government has not focused much attention on computer-aided translation, while the Japanese government has funded significant R&D in this area. Thus, the Japan Information Center of Science & Technology (JICST) began in 1989 to utilize a Japanese/English (J/E) system to translate abstracts into English for their database. The system was recently demonstrated in Tokyo to U.S. Secretary of Commerce Mosbacher.

Because of the basic importance of this technology, the U.S. Department of Commerce sponsored a national conference on J/E computer-aided translation at the National Academy of Sciences in December 1989. Research interest in the U.S. has been revitalized by the event.

The U.S. Task Force is organizing a study team that will visit Japan this year to review the state of the art in Japan and lay the groundwork for specific cooperative projects.

Executive Summary**Contents of this Report**

1. Improve Awareness and Understanding of.....1	
Organizations and Systems Established to	
Improve the Use of Scientific and Technical	
Information (Mutual Education)	
2. Increase the Quantity and Quality of Scientific.....3	
and Technical Information (STI)	
3. Reduce Impediments, If Any, to the Flow of.....6	
Scientific and Technical Information (JRRC)	
4. Increase the Translation of Scientific and.....7	
Technical Information (Translation)	
Appendix A. Present Status of Japanese.....10	
Scientific and Technical Information Centers	
Appendix B. Important Examples of Scientific 15	
and Technical Exchange	
Appendix C. Task Force Participants.....16	
Appendix D. Bibliography.....19	

1. Improve Awareness and Understanding of organizations and systems established to improve the use of scientific and technical information (MUTUAL EDUCATION)

Three activities are in progress: 1) Reports on the use and distribution of Japanese STI; 2) collection of videotapes describing Japanese research facilities; and 3) a conference for American researchers on access to Japanese STI.

(I.) Studies of the Distribution and Use of Japanese Scientific and Technical Information (STI)

A. A survey was made by the Science and Technology Agency (STA) through the Mitsubishi Research Institute (MRI) on present conditions concerning the distribution of Japanese STI to the United States. The results of the survey pointed out the need to improve the accessibility of Japanese technical reports, doctoral dissertations, and technical bulletins/newsletters, and the necessity to enhance conditions of distribution to the United States by increasing directory information and reducing the language barrier between both countries. The survey results are contained in a report titled *Research on the Dissemination of Japanese Scientific & Technical Information*.

B. A follow-up study to Prof. Morita's report, *The Current Status of Science and Technology Grey Literature in Japan*, was made by the National Technical Information Service (NTIS). The study by M. Ruhl was intended to document the experiences of a non-Japanese librarian in Japan seeking information on a specific subject. The report of the study is titled *Identifying and Acquiring Grey Literature in Japan of High-Definition Television (HDTV) for the National Technical Information Service (NTIS)*.

(II.) Video Tapes Showing Laboratories

During the April 1989 meetings in Tokyo, a few English-language videotapes of Japanese research institutions were given to a U.S. delegation member. Such video tapes showing the activities of government and government-funded research facilities are useful because:

- They serve to familiarize researchers with the facilities, capabilities and interests of R&D laboratories.
- They provide researchers, officials and businessmen with a preview of laboratories that they may wish to visit, and prepare them for the visits.

It is recommended that video tapes be exchanged as much as practicable. NTIS will serve as the collection and distribution point for the United States.

(III.) Conference to Promote the Distribution of Japanese Scientific and Technical Information

The purpose of the conference will be to:

- Familiarize information intermediaries and researchers in the U.S. with the kind of Japanese STI that is available, and how it may be obtained.
- Encourage information intermediaries in the U.S. to educate their users as to the availability and importance of Japanese STI.

It is recommended that the Japan Information Center of Science and Technology (JICST) and NTIS mutually agree on the time, location and agenda of the conference.

2. Increase the Quantity and Quality of Scientific and Technical Information (STI)

At the meeting in Tokyo, April 24, 1989, of the Task Force, it was concluded that further investigation of comparable access to STI between both countries was necessary. The Task Force undertook the following first steps:

- Clarify the definition of 'Grey Literature'
- Identify specific opportunities to improve the flow of 'Grey Literature'

With reference to the first listed step above, the Task Force mutually agreed that the term 'Grey Literature' be eliminated and emphasis be placed on exchange of important types of S&T information (see Appendix B). The Task Force agrees that the definition in Annex II of the original cooperative agreement be adopted for the purposes of this exchange as follows:

Scientific and technical reports produced by government agencies or through major government-sponsored or government-supported research programs that are not published in readily available professional literature.

Significant steps taken recently to improve the STI flow include:

a. The National Diet Library (NDL) of Japan established a post in their Acquisition Department in July 1989 to specifically promote the deposit of central and local government publications.

In cooperation with UNESCO, the NDL published the *Directory of Japanese Scientific Periodicals 1988* in April 1989, to increase international awareness of Japanese STI publications.

b. JICST added English bibliographic citations of government reports to their JICST-E file in October 1989. JICST has used a machine-aided translation system to translate about 1,700 bibliographic citations with abstracts from Japanese to English in JFY1989. These citations are being added to the JICST-E file.

As an International MEDLARS Center for Japan, JICST began to add Japanese (translated) MeSH terms to NLM databases to help Japanese customers search MEDLINE, TOXLINE and CANCERLIT in Japanese.

JICST also assisted in a survey NTIS conducted to identify the various types of STI publications in Japan.

c. The National Center for Science Information System (NACSIS) created a new database system on research being conducted at universities and academic societies and associations. This database is now available to the U.S. scientific community through the Library of Congress and the National Science Foundation. Additionally, an electronic mail service was created

through CSNET and through BITNET at George Washington University.

Following the conclusions of the Task Force meeting in 1989, the Japanese side intends to implement the following, beginning in JFY 1990 on April 1, to further improve the flow of STI:

a. Starting JFY90, JICST has an initial annual budget of approximately \$1 million to develop an English bibliographic database of STI such as reports of national institutes or laboratories.

JICST will also develop a database of conference pre-prints.

JICST will prepare directories in English for contact points in laboratories/research institutes and Japanese STI suppliers.

In the U.S., JICST will conduct conferences and exhibits at professional meetings to promote distribution and utilization of databases and directories such as those mentioned above.

JICST is making efforts to enlarge the database of the English version of Japanese literature to be provided through JOIS and STN International.

b. The National Center for Science Information System (NACSIS) will promote the development of databases of abstracts on grant-in-aid research sponsored by the Ministry of Education, Science and Culture, theses submitted to universities in Japan, and papers presented to conferences and meetings of academic societies and associations. NACSIS will also promote

development of databases of academic papers issued at Japanese universities.

The U.S. side expressed deep appreciation for the Japanese government's ongoing commitment to this effort, and for the concrete evidence of that commitment in the form of approximately \$1 million in new budget authority for JICST information activities.

3. Reduce Impediments, If Any, to the Flow of Scientific and Technical Information (JRRC)

The current focus of this activity is on the establishment of the Japan Reprographic Right Center (JRRC). Preparation for the establishment of the JRRC is in progress by a group of promoters composed of representatives of authors' associations, publishers' associations and academic societies as well as persons of learning and experience.

The group of promoters is now planning concrete activities of the JRRC, and is preparing for various regulations concerning the composition of the Center, the method of entrusting the JRRC with rights, and the methods of collection and distribution of license fees. The Agency for Cultural Affairs will give guidance from the legal and practical point of view.

There is a need for the JRRC to begin its activities by mutual agreement among authors, publishers and users. The group of promoters is making an effort to obtain mutual agreement among them.

The Japanese government continues to encourage the establishment of a private Japan Reprographic Right Center (JRRC) similar to the U.S. Copyright Clearance Center.

In an address to the Task Force on April 3, 1990, Eamon Fennessy, President of the U.S. Copyright Clearance Center, indicated that the CCC began its operation in 1978 providing service primarily for scientific, technical and medical literature. This was done to simplify the task of starting a reprographic rights organization.

Again, the U.S. side of the Task Force strongly urges early establishment of the JRRC.

4. Increase the Translation of Scientific and Technical Information (Translation)

(I.) Translation Rights Permission

In order for U.S. users to access Japanese literature, it is necessary for documents to be translated into English. At present there is no mechanism to facilitate negotiation of translation rights, although those negotiations are done through many different channels between American users and Japanese authors and publishers.

The Joint Task Force agrees that it is desirable to institute a mechanism to improve this situation. The Task Force recommends that the U.S. and Japan study cooperatively some practicable mechanism which facilitates communication of translation rights negotiation between concerned parties. For this purpose, the Task Force agreed to develop a plan to enhance the dissemination of STI beyond language barriers and to minimize the impediment represented by language barriers.

(II.) Machine-Aided Translation

In order to expedite the flow of Japanese scientific and technical information to U.S. scientists and engineers, the United States and Japan are focusing on a cooperative program on machine-aided translation. Although in the past the U.S.

effort has been primarily confined to European languages, the Japanese have been extremely active in this field, and have already developed several working systems for the Japanese-English pair. The most extensive government effort is the J/E system of the Japan Information Center of Science and Technology (JICST), which is designed specifically to expedite the translation of scientific and technical abstracts for the JICST-E database. This system was demonstrated recently to U.S. Secretary of Commerce Mosbacher during his visit to Japan. The JICST J/E system has already been used to prepare approximately 1,700 English-language abstracts for the JICST-E electronic database during JFY 1989. It is expected that the system will be improved, based on the JFY 1989 experience, and result in considerable increases in computer-translated output during JFY 1990.

Meanwhile, in the United States, a national conference on computer-aided translation was held at the National Academy of Sciences in December 1989, under the sponsorship of the U.S. Department of Commerce. The conference was designed to inform the U.S. audience of Japan's advances in Japanese-to-English machine translation. Professor Nagao of Kyoto University, an internationally recognized expert from Japan, was the keynote speaker. The U.S. community appears to have been energized by this conference.

Since then, the U.S. Task Force has held several planning meetings to organize a study team that will visit Japan in 1990 to review Japanese accomplishments and lay the groundwork for specific cooperative projects. The study team will evaluate JICST's J/E system, as well as other systems, with respect to: their abilities to assist U.S. users who are not familiar with the Japanese language; their fidelity to technical intent and the idiomatic accuracy of their outputs; and such considerations as expandability to popular U.S. PC's and workstations, adaptation of OCR (optical

character readers of the original Japanese text) devices, and accessibility via electronic communications systems. The Task Force expressed appreciation to JICST and Professor Nagao for their willingness to assist the study team in selecting the key sites and experts and arranging the visits.

Candidates for the study team's members and leader are being identified. The U.S. Task Force is aiming for the study team to be comprised of representatives from all expert areas involved, including fundamental research, linguistics and end use. In the long term, team findings are expected to benefit two principal groups: Scientists and engineers who are conducting research on computer-assisted-translation; and users whose research will be expedited by access to Japanese STI. Administration of the visit will be carried out through the National Science Foundation's Japanese Technology Evaluation Center (JTEC) program. The study team will visit Japan in 1990.

Appendix A

Present Status of Scientific and Technical Information Centers

Scientific and technical information centers in Japan may be classified by their activities in the following way:

(1) Comprehensive Information Centers

Two government organizations collect, process and provide the information for all fields of science and technology: The Japan Information Center of Science and Technology (JICST, under the Science and Technology Agency) and the National Center for Science Information System (NACSIS, under the Ministry of Education, Science and Culture).

JICST, a central organization for scientific and technical information services in Japan, has been compiling and providing scientific and technical information from all over the world (740,000 citations in FY 1988 of which 360,000 are domestic sources). Recently, it has started providing factual databases, with 7 databases at present. JICST is financed by two income sources; namely, governmental support and service fees.

NACSIS, a national inter-university research institute, was established in 1986 as the nucleus organization of scholarly information and provides the services of compilation, processing and dissemination of scholarly information including scientific and technological fields mainly for academic researchers. It is also conducting research and development concerning science information systems. NACSIS databases include union lists of approximately three million pieces of data from university libraries.

(2) Specialized Information Centers

The centers collect, process and provide scientific and technical information in particular fields. Most of them are governmental agencies in charge of those particular fields or non-profit corporations established under the supervision of the authorities. The following are the main organizations:

- Japan Patent Information Organization (JAPIO):
patent information
- Small and Medium Enterprise Information Center,
Japan Small Business Corporation: technical
information for small and medium enterprises
- Japan Pharmaceutical Information Center (JAPIC):
pharmaceutical information
- Japan Medical Abstracts Society (JMAS): medical
information; JICST provides its information through
online service system; a voluntary organization.
- Agriculture, Forestry & Fisheries Research
Information Center, The Ministry of Agriculture,
Forestry and Fisheries: agriculture, forestry, and
fisheries information
- International Medical Information Center (IMIC):
medical information
- The Japan Iron and Steel Technical Information
Center, The Iron and Steel Institute of Japan:
iron and steel information

- Japan Association for International Chemical Information (JAICI): chemical information
- Japan Industrial Technology Association (JITA): MITI's patent license information

(3) Collection, Storage and Dissemination Centers

Libraries and JICST are the major organizations which collect and store primary information. According to the National Diet Library (NDL) Law, all unclassified publications (including grey literature) published in Japan are to be deposited in the NDL. The NDL develops databases of literature collected and provides it on magnetic tapes or CD-ROMs. It is also provided to public libraries using an online system.

This database is called JAPAN/MARC. It has about 900,000 records now and adds 60,000 records annually. 12.5% of them (about 110,000 records) are scientific and technical books, 22% of them are government publications. Scientific and technical records have major proportion of government publications. The Government Publications Codes are assigned to the records of government publications.

University libraries are the major centers for the collection and storage of scholarly materials both domestic and foreign. Most of these libraries and JICST provide a copy service upon request. According to the Copyright Law, only one copy of a part of a publication is permissible.

NACSIS has developed a database for holding information of scientific journals and monographs collected in the

libraries of universities and major domestic research institutions through their mutual cooperation and also provides those retrieval services. Each university library collects, stores and provides those materials.

(4) Database Distributors

Besides public organizations such as JICST, NACSIS and JAPIO, some private companies provide online services of scientific and technical information. Some others provide online service to domestic users as representatives for foreign online vendors such as DIALOG and BRS.

(5) Information Analysis Organizations

Demand for information with deeper analysis has been increasing and many information organizations provide such value-added information services. Besides private think tanks and information brokers, organizations such as Japan Small Business Corporation and regional information centers for small and medium enterprises provide very detailed technical and managerial information.

(6) Clearing Houses

JICST provides online service of on-going research projects carried out at public research institutes in Japan. The publication corresponding to this database, "Current Science and Technology Research in Japan" is also provided in print.

JICST also has constructed a database for government reports, such as survey reports or research reports by

municipal corporations and research reports of the projects funded by governmental organizations.

NACSIS, on the other hand, develops a database of reported synopses of research projects conducted through research grant-in-aids by the Ministry of Education, Science and Culture and provides it online.

Appendix B

**Important Examples of Scientific and Technical
Information for Exchange**

- Reports of national institutes or laboratories
- National government publications
- Official government documents released to foreign organizations
- Conference papers and proceedings
- University publications including theses and dissertations
- Publications of professional societies and associations
- Access to database (bibliographic, numeric, graphic, full text) services

Appendix C

U.S. - JAPAN TASK FORCE
ON
SCIENTIFIC AND TECHNICAL INFORMATION
APRIL 3, 1990

JAPANESE PARTICIPANT LIST

Mr. Yasutsugu Takata (Co-chair)	Deputy Director-General Science and Technology Promotion Bureau Science and Technology Agency
Mr. Chiyohiko Hakoishi	Director Science and Technology Information Division Science and Technology Promotion Bureau Science and Technology Agency
Mr. Hiroshi Nagano	Director International Affairs Division Science and Technology Promotion Bureau Science and Technology Agency
Mr. Hiroshi Kimura	Science and Technology Information Division Science and Technology Promotion Bureau Science and Technology Agency
Mr. Yubun Narita	Director Scientific Affairs Division Ministry of Foreign Affairs
Mr. Junji Yamamoto	Coordinator for International Science Program International and Science Bureau Ministry of Education, Science and Culture (MONBUSHO)
Mr. Teruhide Fujita	Director International Research Division Research Council Secretariat Ministry of Agriculture, Forestry and Fisheries

JAPANESE PARTICIPANT LIST
(Cont.)

Mr. Kaoru Honjo	Director for Planning of Regional Technology Agency of Industrial Science and Technology Ministry of International Trade and Industry
Mr. Keikou Terui	Deputy Director International Research and Development Cooperation Division Agency of Industrial Science and Technology Ministry of International Trade and Industry
Mr. Kazuhiko Fuji	Deputy Director General Coordination Division Agency of Industrial Science and Technology Ministry of International Trade and Industry
Mr. Takashi Kisaka	Counsellor Embassy of Japan
Mr. Masakazu Toshikage	First Secretary Embassy of Japan

U.S. PARTICIPANT LIST

Clark, J.
Department of Commerce/NTIS
Co-Chair

Billingsley, R.
Defense Technical Information Center

Carter, C.
National Academy of Sciences

Feinstein, T.
Department of Commerce/NTIS

Fennessy, E.
Copyright Clearance Center

Khan, R.
Department of Commerce/NTIS

Kusuda, T.
Department of Commerce

Rudin, E.
National Science Foundation

Shonyo, D.
Department of Commerce/NTIS

Staheli, L.
Department of State

Wright, G.
Office of Naval Research

Appendix D

Bibliography

Agreement Between the Government of the United States of America and the Government of Japan on Cooperation in Research and Development in Science and Technology. Toronto, Canada. June 20, 1988.

Morita, I.T., Current Status of Science and Technology: Grey Literature in Japan (PB88-227780). Springfield, Virginia:National Technical Information Service, 1988.

Report of the December 1989 Japanese-to-English Machine Translation Conference (in preparation). National Academy of Sciences.

Research on the Dissemination of Japanese Scientific and Technical Information. Mitsubishi Research Institute and Science and Technology Agency. March 1989

Ruhl, M.J., Identifying and Acquiring Grey Literature in Japan on High-Definition Television (HDTV) for the National Technical Information Service (PB90-135302). Springfield, Virginia:National Technical Information Service:1989.

Task Force Report on Information Access. Tokyo, Japan. April 24, 1989.

Report to Congress

Current Status
of
Japanese to English Machine Translation

July 1988

Report to Congress

Current Status
of
Japanese to English Machine Translation

July 1988

C. William Verity	Secretary of Commerce
Robert Ortner	Under Secretary of Commerce for Economic Affairs
Barry C. Beringer	Associate Under Secretary of Commerce for Economic Affairs
Sabina G. Javits	Office of Japanese Technical Literature
Tamami Kusuda	Office of Japanese Technical Literature

Introduction

Everyone would agree that the biggest obstacle to the effective use of Japanese technical literature and/or information by American engineers and scientists is the language barrier. The Japanese language is little known outside its native land, one of the most difficult of the world's major languages for a Westerner in which to gain reasonable fluency. In order to become literate in Japanese technical publications, one must devote a considerable amount of time and concentration to study the language, both of which most engineers and scientists in the United States can ill-afford. It is expected that most Americans will continue to read their Japanese literature in English translation for years to come. There is, however, a shortage of qualified Japanese-to-English technical translators everywhere, inclusive of Japan, and the costs of translation tend to be considerably higher than the costs of translating comparable materials from European languages. (It easily ends up costing as high as \$100 per page). In addition, the accuracy and "correctness" of translations tends to be low, because most translators are usually not proficient in a variety of advanced technical disciplines.

Under these conditions, machine translation appears to offer some hope for a solution to this problem. The development of a fast and accurate machine translation (MT) system (or computer aided translation system) would increase the availability of Japanese technical literature in the United States, and expedite the dissemination of Japanese information on scientific discoveries and technical innovations.

Machine Translation Process:

The internal workings of any MT algorithms are extremely complex, vary according to each individual program, and will not be covered in this report. In addition to a common words dictionary and a special technical terminology dictionary, most Japanese to English MT programs have the following elements: (1) Input of a sentence after a pre-editing process, (2) Morphological analysis of a string of Japanese characters to form meaningful words, (3) Syntactic analysis of these words to form a syntactic dependency structure, (4) Semantic analysis of the syntactic dependency structure to form a semantic dependency structure for Japanese, (5) Transformation of the Japanese semantic dependency structure into that of English, (6) Syntactic generation of an English phrase structure (tree structure), (7) Morphological synthesis of an English sentence with appropriate tense, articles, preposition and post positions, and number, etc., (This is not easy, because Japanese lacks the concept of articles and in many instances it is sloppy in plurality distinction), and finally (8) Output of a translated English sentence followed by a post-editing process. In addition to this basic structure, good MT programs should include algorithms on contextual analysis or intersentential relationships.

The process described herein is the so called the "transfer method", which is most popular among current MT systems. There is another machine translation system that uses the "interlingua method", which will be described later.

Although considerable advances in MT applied to European languages have been made in recent years, even the current best MT program is still far from the stage that it can replace human translators. According to W. John Hutchins, author of "Machine Translation: Past, Present, Future" (Chichester: Horwood, 1986.), no current operational or established MT system can produce good quality output without either placing restrictions on input or involving human assistance before, during and after translation processes. Since the present MT system makes simple errors which no human translator would possibly make, it is not surprising that in general experienced translators have low opinions of MT systems, including the most advanced systems.

US Activities in Japanese to English MT

In spite of strong and expressed needs, US government efforts to develop Japanese-English machine translation are limited to a project under way by the Air Force Foreign Technology Division (FTD) and the Intelligence Community. This project was formally initiated in 1985 following several years of planning and consideration of various machine translation options available in the United States and Japan.

The Intelligence Community chose to support research and development of a Japanese-English Systran System (JESS), based in part on the successful development of a Russian-English Systran system. This Russian system played a vital role during the Soyuz-Apollo mission when large volumes of Russian technical documents had to be rapidly translated into English. In addition, Systran has been implemented for the translation of French, German, Italian,

Spanish, Portuguese, and Dutch. At present it is the most widely-used MT system in the world by governments and industry including the US Air Force, NASA, the European Communities (CEC), Xerox Corporation, General Motors of Canada, etc.. Unlike other contemporary MT systems, which rely heavily on high level programming languages such as LISP, C, and Prolog, Systran is written primarily in a proprietary macro language designed specifically for handling linguistic tasks. In addition, parts of some utility and system control programs are written in assembler. It is considered to be much faster than other systems. An additional factor in Systran's favor, which led US government support for its Japanese-English translation project, was the fact that a Japanese firm called "IONA" had already invested considerable resources in developing the basic software needed for a Japanese to English system in the general Systran framework, which included other language pairs. This work is being conducted under a contract with a US company called Latsec, Inc., the developer of Russian-English Systran. The US government has a complicated licensing agreement with IONA and WTC (its government branch is LATSEC) to have access to work on JESS with the proviso that US government use of the system outside the Air Force, after the system reaches production status, is limited to the Foreign Broadcast Information Service (FBIS), the Office of Patent and Trademarks (OPT), and the National Technical Information Service (NTIS).

Since the US Government-sponsored work on JESS officially started in early 1986, Latsec's work has been primarily directed at expanding the dictionary and acquiring new input hardware with

limited linguistic work. The Intelligence Community hopes to have a prototype available in late 1988 to mid 1989.

In addition to this US Government sponsored project, MT research is being conducted at the Carnegie Melon University's (CMU) Center for Machine Translation under the direction of Professor Jaime Carbonell and Masaru Tomita. They have developed an advanced multilingual system called the "Universal Parser", which is an interlingual system. Unlike the transfer method mentioned earlier, the interlingua MT is designed to perform multi-lingual translation through a pivot language. Any pair of translated languages must go through this pivot as opposed to direct transfer from one language to another. The Japanese to English system at CMU is a part of this interlingua system, which currently handles Japanese, German and English for simple medical diagnostic related sentences with limited vocabulary and grammar. The system has also been tested for use in speech translation. These CMU projects are supported by the National Science Foundation and Japanese corporations.

Other Japanese to English MT-related activities have been reported by the Hunter College SUNY, Martin Marietta Corporation, and at the Courant Institute of Mathematical Sciences at New York University. Details of these efforts were not available at the time of this report.

Japanese Activities on MT

In contrast to the rather quiescent state of US efforts, the Japanese government views Japanese-English machine translation as an essential part of its future information-based society. Beginning in

1982, ample funding has been provided for the MU (and subsequently for MU-II) projects to develop a suitable MT system for translating JICST (Japanese Information Center for Science and Technology) English Data Base abstracts. JICST English Data Base is one of the most comprehensive bibliographic files covering approximately 5000 journals as well as technical reports in every field of science and technology published in Japan since 1985. Its English abstracts are currently prepared by a large contingent of carefully chosen technical consultants, yet they are usually behind as long as six month after the publication of original articles, a very serious drawback for this day of rapid technical advances. The Japanese government considers the JICST English Data Base to be one of crucial windows of its technology internationalization policies and is very anxious to expedite the availability of most up-to-date abstracts.

The MU project was a well developed and typically Japanese style cooperative research and development effort among industry(San), government (Kan) and university (Gaku). The project was led by Professor Makoto Nagao of Kyoto University, a father of Japanese MT research. Although the funding from the Science and Technology Agency through MITI (Ministry of International Trade and Industry) was originally 600 million yen during 1982-86, the real cost of this venture was much larger. Ten computer manufacturing companies participated in this MU project by contributing researchers to Prof. Nagao's laboratory. These company researchers later played central roles in the development of their commercial MT systems, most of which were demonstrated during the MT Summit Conference held in

Hakone, Japan in September, 1987. A brief summary of these systems is presented in the attached table. According to Prof. Nagao, currently over twenty companies in Japan are engaged in the development of commercial Japanese-English machine translation systems.

In addition, the Japanese government supports several MT related systems, such as Fifth Generation Computer System (ICOT), Electronic Dictionary Research projects (EDR, IPA), and CONTRAST of the Electro Technical Laboratory of MITI/AIST (Agency for Industrial Science and Technology). The Ministry of Posts and Telecommunication supports Automated Translation Telephone Research (ATR). Recently MITI initiated the CICC (Center for International Cultural Corporation) project to assist four neighboring countries in developing an inter-lingua MT program for multi-lingual translation of Chinese, Thai, Malaysian and Indonesian. MITI is expected to allocate 6 billion yen (\$45 million) to develop a truly multi-lingual MT system centered around Japanese. When completed, this advanced system could also be used to enhance Japanese-English translation technology.

Most commercial Japanese systems are produced in an integrated system, or workstation, comprised of a company specific mini-compute, MT software, word processor, and input/output devices (inclusive of an optical character reader, or OCR for some systems), and are not available for software alone. An exception is one of the most commercialized systems called the Bravice system, originally developed by Weidner of the United States and later sold to a Japanese company. This is the only Japanese to English MT system

available for personal computers (unfortunately not for the popular US made PC's but for popular Japanese PC's such as NEC 9800, IBM 5550, Toshiba, etc.). The Department of Computer Science at New York University is currently evaluating the Bravice system for US use with the NEC PC-9801, which is available in the USA).

Most commercial Japanese MT systems have a basic dictionary of 50,000 words supplemented by an additional dictionary of 200,000 to 250,000 technical words. The translation speed is dependent upon the computer system, ranging from 60,000 words/hr of the main frame computer systems to 3000~5000 words/hr of the microcomputer based systems. The accuracy of their MT systems is generally claimed to be in the order of 70~80 % depending upon the complexity of the source document. With good pre-editing, some tasks showed an accuracy of over 90%.

Their MT systems have conversational editors and the ability to show on the split screen the source (Japanese) and the target (English) texts displayed side by side, making it easy to apply pre- and post editing to achieve the desired results. They are definitely a boon to professional translators, since the majority of technical terminologies are correctly and consistently translated by their technical dictionaries.

All Japanese commercial MT systems use the transfer method, except for PIVOT of NEC and ATLAS-II of Fujitsu, which employ the interlingua method (It should be noted that definitions of these classifications differ. While ATLAS-II claims to use an interlingua model, its published system descriptions refer to a transfer component). Prof. Nagao considers the transfer method to be more

suitable for conventional technical papers, business letters, manuals, etc, while the interlingual approach is more appropriate for cases in which the source text is fragmented, incomplete, erroneous and non-grammatical requiring in-depth linguistic analysis. There are disagreements among Japanese MT developers as to the merit of the interlingua approach, which appears to be more challenging and suitable for multi-lingual translation, yet is still considered somewhat over-theoretical.

Although many Japanese commercial systems appear impressive in demonstration, the true state of their accuracy is best described by Prof. M. Nagao as follows: "The companies selling machine translation systems are now being criticized by their customers for the poor quality of their translations and are agonizing over the difficulty of making significant improvements in their systems. It is an illusion for any MT system user to expect high quality results, unless the original text is carefully written to avoid ambiguity and misunderstanding." Prof. Nagao believes that an MT system can become a viable undertaking if one limits its application to

1. Daily news
2. User's manual for export/import merchandise
3. Abstracts
4. Drafts of conference memos
5. Drafts of stylized business letters

European Activities for Japanese to English MT

In England, there are two Japanese-English MT activities worth noting. The first is the AIDTRANS system of the Center for Japanese

Studies at Sheffield, based on a highly sophisticated and detailed analysis of Japanese grammar and the Integrated Dictionary System (IDS). The second system is the UMIST English-Japanese NTRAN system designed to permit technical authors of computer software manuals to read and write such texts in Japanese. As yet none of these English system are available for evaluation.

The European Community (CEC) has contracts with Systran and Fujitsu for pilot projects to produce Japanese-English translations to the community.

Conclusions and Recommendations

Unlike Japan and the CEC countries, MT research in the United States has lagged due primarily to the devastating effects of the famous ALPAC (Automatic Language Processing Advisory Committee) report published in 1965. It is important to mention here that the MT research started in the United States before the 1950's. In 1954, IBM and Georgetown University already developed an assembly language machine translation system, a forerunner of Systran. The National Science Foundation (NSF) spent nearly \$20 million between 1955 and 65 supporting various MT projects. The progress was very slow, and after careful and exhaustive studies on then active MT research activities in the United States, the ALPAC report conclusively recommended the termination of government support for MT activities by stating that the cost of developing viable MT systems to provide comparable translation done by human translators would be prohibitive (still a prevailing attitude within some circles). Since then the US Government support on MT has

been confined to US Air Force project for the Systran system, which was very effective in achieving over 90% accuracy for Russian to English translation of technical material. However, the Systran's work on Japanese to English MT system(JESS) has not been significant yet.

On the other hand, the Japanese have made impressive progress in developing commercial MT systems, which appear adequate for translating so called "minimum Japanese" into "minimum English". In other words, the translated results of technical documents prepared with rigid formats and grammar, such as JICST abstracts, user's manuals and service manuals for machinery and electronic instruments, are easily understood by readers familiar with the subject matter. There is no doubt that the Japanese will keep improving their systems so that acceptable quality MT systems will be made available at low cost.

Current Japanese MT systems, however, are still useless for the majority of American engineers and scientists who are unable to input Japanese texts onto the keyboard. An optical character reading (OCR) system to read Japanese text including charts, figures and tables is mandatory. A minimum of four Japanese companies, Fuji Electric, Ricoh, Sanyo and Oki Electric are currently marketing OCR. Most have the capability to read different size fonts in vertical as well as horizontal sentences. Some experimental systems can even read handwritten Japanese text. The speed of character recognition varies from 5 characters per second to 30 characters per second depending upon the nature of font with a claimed recognition accuracy of over 99%. The prices of their OCR's are, however, rather

high ranging from Ricoh's \$8000 to Fuji's \$37,000. Oki is said to be planning to market an integrated OCR-MT system by Sept. 1988.

It is recommended that these Japanese accomplishments be evaluated critically by competent US experts on MT to see if some of their commercial systems can be adopted or made use of in the US. There is a crucial need to expedite the use of MT by US engineers and scientists who want quick assessments of the technical content of Japanese literature without going through the costly, time-consuming, and not necessarily useful process of full translation.

Japanese to English Machine Translation Systems In Japan
(Information available as of Sept. 1987)

Name /Company	Computer/ Language	Dictionaries Common/Technical	Speed Words/hour	OCR Connection
Atlas-II Fujitsu	FACOM /C	50,000/250,000	60,000	Yes
HIGHCATS Hitachi	HITAC/PL1	50,000/250,000	60,000	No
TAURAS Toshiba	UNIX/C	50,000/50,000	7,000	No
SWP-7800 Sanyo	IRMX/C	50,000/55,000	3,500	Yes
PENSEE Oki	UNIX/C	60,000/40,000	4,000	Yes
PIVOT NEC	ACOS/	40,000		No
MELTRAN Mitsubishi	SIMPOS/ESP	50,000/30,000	5,000	No
LAMB Cannon	Symbolics/LISP	2,000	1,000	No
PAROLE Matsushita	General/Prolog	5,000		
BRAVICE Bravice	Japanese PC's	70,000/30,000	2,000	No
MU Kyoto Univ.	FACOM/LISP	300,000	4,000	No
Systran IONA	FACOM/ Proprietary Assembler	40,000/(over 300,000)	----	----

Bibliography

Following materials were useful in the preparation of this paper.

1. Appendix 5, "Machine Translation", Report to Congress "Activities of the Federal Government to Collect, Abstract, Translate and Distribute Declassified Japanese Scientific and Technical Information 1987-1988", PB88-194816, Office of Japanese Technical Literature, US Department of Commerce, 1988
2. MT Summit, Manuscript and Program, Japan Electronic Industry Development Association, September 17-19, 1987
3. David B. Shonyo and Janet H. Geffner, "The Supply and Use of Japanese Technical Information in the United States", National Technical Information Service, 1988
4. Program of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, June 12-14, 1988, Center for Machine Translation, Carnegie Mellon University
5. Carol Pearson "Moving Beyond Vodka and Rotten Meat" Carnegie Mellon Magazine, Vol.6 No.3, Spring, 1988, pp 14-19.
6. Private conversations with Mr. Tatsuo Ashizaki, Chief Machine Translation Section, Japan Information Center for Science and Technology
7. Private conversations with Ms. C.C. Walrad of Systran Corporation.

Those who wish to obtain further information on this report should contact

Dr. T. Kusuda
Office of Japanese Technical Literature
US Department of Commerce
Room 4833, HCHB
14th and Constitution Ave.
Washington, DC 20230

Telephone (202) 377-1287

JAPAN'S STATE-OF-THE-ART IN MACHINE TRANSLATION TECHNOLOGY

Emily B. Rudin
International Science Analyst

La Cour de Justice envisage la création d'un cinquième poste d'avocat général.

The yard of justice is considering the creation of a fifth general avocado post.
(First attempt by a machine to translate above French sentence into English)

The Court of Justice is envisaging the creation of a fifth advocate-general's post.
(Same sentence after post-editing by a human translator)

Introduction

Sometime before the end of 1990, a U.S.-Japan Task Force on Scientific and Technical Information will launch a study of Japanese efforts in machine translation technology. Supported in part by the National Science Foundation (NSF), this project is intriguing, in that it represents a convergence of the respective interests of broad national policy and narrowly defined basic research. It will be conducted by a select team of experts who will travel to Japan not only to evaluate the status of research and development in this field, but also to lay the groundwork for future cooperative research between the United States and Japan.

The project will contribute to the knowledge base in several important ways. *First*, it will provide an up-to-date assessment of Japanese capabilities in this field. *Second*, it will identify strategic opportunities, through bilateral collaboration, that will help to advance the state-of-the-art for mutual benefit. *Third*, and perhaps most significantly, it will accelerate progress in machine translation technology for the long term. Advanced capabilities in this area will greatly enhance the flow of scientific and technical information between the two countries, reducing the language barrier and making possible the transfer of research results much more quickly and comprehensively than ever before.

The 1990s have been heralded as the Decade of the Pacific. Given the premise that Japan will continue to be the key scientific and technological player in that unfolding scenario, real-time access to the unclassified and nonproprietary universe of

Japanese R&D will be essential if the United States is to capitalize on cutting-edge scientific discoveries and technological innovations emerging in Japan. Toward that end, machine translation of Japanese research reports will have a crucial role to play, in this decade and beyond.

[Editor's Note: For purposes of this article, the term *machine translation* (often referred to as MT) is defined by the technical community as translation which is generated by a computer, with or without human intervention. No extant machine translation system in the world is yet capable of total self-sufficiency.

[A common misperception is that the term *machine-aided translation* can be used interchangeably with MT. According to technical definition, machine-aided translation is *not* generated by a computer; rather, it provides leverage for human translators with such computerized assets as terminology banks, grammar checkers, and word processing.

[Both of these terms are subsumed under the umbrella designation, *machine(-assisted) translation*, or M(A)T. While the scope of the study team is focused primarily on MT, the Task Force expects that machine aids to translation will also be investigated. Throughout this article, exact terms will be specified wherever appropriate.]

Background: Increasing U.S. Access to Japanese Scientific and Technical Information

To understand the importance of machine translation at this juncture, a brief look at Japan's increasing R&D strengths, along with an overview of recent U.S. Government steps toward improving access to Japan's research results, might be useful.

Powerhouse of Scientific and Technological Output. The Japanese are reinforcing their basic and applied research capabilities by applying revenue from their favorable balance of trade with other countries. These investments have begun to produce a highly competitive economic profile. A random sample of science and technology indicators provides an idea of the magnitude of Japan's productivity and priority with respect to R&D:

- In 1986, Japan contributed 13 percent of the world's literature in engineering and technology—a notable increase in its world share from 9 percent in 1981.
- In 1985, Japan ranked fourth worldwide in the number of scientific papers published. By 1987, it had accelerated to second place, behind the United States.
- By current estimates, Japan publishes over 10,000 scientific and technical journals annually, as well as an array of books, conference proceedings, reports, and monographs.
- The Japanese government's R&D expenditure has expanded by approximately 25 percent since 1984—even though the total government budget has increased by only 1 percent per year.

Tapping Japan's R&D Strengths. During the 1980s, increasing U.S. attention was devoted to making this growing pool of Japanese literature more available to American researchers. The Stevenson-Wydler Technology Innovation Act of 1980 took a major step in this direction, with its emphasis on facilitating the transfer of research data from Government to the private sector. In 1986, the Congress passed an amendment to the Act, entitled the Japanese Technical Literature Act. This legislation charged the National Technical Information Service (NTIS), and other offices within the Department of Commerce, to:

- a. Monitor Japanese technical activities and developments.
- b. Consult with businesses, professional societies, and libraries in the United States regarding their needs for information on Japanese developments in technology and engineering.
- c. Acquire and translate selected Japanese technical reports and documents that may be of value to agencies and departments of the Federal Government, and to businesses and researchers in the United States.
- d. Coordinate with other agencies and departments of the Federal Government to identify significant gaps and avoid duplication in efforts by the Federal government to acquire, translate, index, and disseminate Japanese technical information.

More broadly conceived, but with specific implications for the procurement of Japanese scientific and technical literature, was Presidential Executive Order 12591, "Facilitating Access to Science and Technology," issued in April 1987. This mandate reinforced the intent of the Stevenson-Wydler Act, particularly with respect to international science and technology, by directing Federal agencies "....to ensure that the United States benefits from and fully exploits scientific research and technology developed abroad...." Agencies (including NSF) have been working—both individually and in combination—to implement activities cited in the Order.

In 1988, a Joint Task Force on Scientific and Technical Information was established under the auspices of the U.S.-Japan Agreement on Science and Technology, signed in June of that year by President Reagan and Prime Minister Noboru Takeshita. A major part of the Task Force's mission is to "increase the quantity and quality of scientific and technical information and its translation." The chairman of the U.S. side of the Task Force is Joseph E. Clark, deputy director of NTIS. The Japanese side is chaired by Yasutsugu Takata, deputy director-general of the Science and Technology Promotion Bureau of the Japanese government's Science and Technology Agency.

Machine Translation: On the Frontier

From its inception, the Task Force identified machine translation as an area for special focus. [Editor's Note: References are made below to technical terms associated with machine translation technology. Details on MT and its complexities are provided at the end of this article.]

The Appeal of MT. Translation is a very powerful, cost-effective computer application that can significantly boost scientific and technical exchange between the United States and Japan. An "intelligent" MT system can augment a human translator's efforts by many orders of magnitude: one has only to compare the human rate of thousands of words per day to the computer-generated output of 5,000 to 60,000 words per hour (depending on the use of microcomputer- or mainframe-based systems, respectively). The cost differential is likewise great: human translators may charge upwards of \$100 per page to convert text from Japanese to English; a commercial MT system in the current generation sells for well under \$100,000. For most Japanese commercial systems, there is an accuracy rate of 70 to 80 percent, and as high as 90 percent with good pre-editing by a human translator. These systems feature a basic dictionary of 50,000 words, supplemented by a technical dictionary of 200,000 to 250,000 terms.

Although a machine-generated translation may be in rough prose and require further hands-on editing, the accelerated rate and volume of translation, as well as the proportionately lower costs, make possible more efficient filtering of the material for its potential utility. One corporate U.S. expert on Japanese scientific information explained the value of this enhancement thus: "There is tremendous demand for foreign technical literature, but someone has to take the lead in seeding that demand. Researchers don't know it is important until they see it. They don't see it until someone not only develops it, but screens it for the useful parts."

Japanese Efforts in MT Research. Since the mid-1950s, the Japanese have been extremely active in the field. The impetus for their research can be partially ascribed to their hunger for information from abroad. In large measure, however, it can be interpreted as a spinoff of their heavy exportation of high-technology products,

requiring massive documentation in users' languages and high expenditures for translation (in 1987 alone, Japan spent \$6.8 billion for translation services). As a result, the Japanese have developed first-generation computer systems to facilitate English-to-Japanese and Japanese-to-English translation, and have produced a number of working systems for the Japanese-English language pair. In 1984, indigenously developed MT systems were introduced commercially in Japan. The use of these systems is increasing, especially in Japan and in Europe. Japanese R&D promises to expand system capabilities even further, enhanced by theoretical work on a universal translator "language," called an *interlingua*. There is also work on translation of voice input. A reasonable outlook for the not-too-distant future points toward low-cost systems which will be able to produce Japanese-to-English translations of a quality acceptable for scientists and engineers to conduct keyword searches or to rapidly assess technical content.

Japan's R&D efforts in MT are largely concentrated in private industry, although some work is also conducted in universities and in other research centers. In the governmental sector, one of the most extensive initiatives is the Japanese-to-English system of the Japan Information Center of Science and Technology (JICST), which was designed to expedite the translation of scientific and technical abstracts for the JICST-E (English) electronic database. In March 1990, this system was demonstrated to U.S. Secretary of Commerce Robert Mosbacher during his visit to Japan. JICST's Japanese-to-English MT system has already been pressed into full service: during Japanese Fiscal Year 1989, it prepared approximately 1,700 English-language abstracts for the JICST-E database. For JFY 1990, JICST officials predict that the system will be improved and produce even more abstracts. [Editor's Note: The JICST-E database is available in the United States through NTIS, by subscription. For further information, contact: Barbara Payne, National Technical Information Service, 5285 Port Royal Road, Room 306F, Springfield, Virginia 22161; telephone: (703) 487-4826; FAX: (703) 321-8199.]

Other MT projects supported by the Japanese government include development of the Fifth Generation Computer System (ICOT); research on electronic dictionaries and related topics at the

Electro-Technical Laboratory of the Agency for Industrial Science and Technology, under MITI (Ministry of International Trade and Industry); and research on automated interpreting telephony by Advanced Telecommunications Research Institute International (ATR), sponsored in part by the Japan Key Technology Center. Several years ago, MITI established the Center for International Cultural Cooperation, with an allocation of approximately \$45 million. This Center is intended to assist the region in developing an MT system that uses an interlingua for multilingual translation of Chinese, Thai, Malaysian, and Indonesian—with Japanese as the central reference point. When this advanced system is fully developed, it could be applied to enhance technology for Japanese-to-English translation.

U.S. Efforts in MT Research. In contrast to the Japanese, U.S. research in machine translation technology has been fairly quiescent. It was not always thus. MT research had its genesis here before 1950. By 1954, IBM and Georgetown University had jointly developed a pilot Russian-English MT system. Between October 1954 and May 1965, NSF expended nearly \$6.6 million on support of MT research, most of it confined to European languages. Plagued by slow progress, MT projects received a death sentence in 1966 with the publication of a report by the Automatic Language Processing Advisory Committee (ALPAC) of the National Academy of Sciences/National Research Council. After an intensive survey of MT research in the United States, ALPAC recommended that governmental support of MT projects be terminated. The committee's justification hinged on the conviction that the cost of developing viable MT systems that could equal or exceed the quality of human translations would be prohibitive.

For many years, the ALPAC report devastated U.S. research in this field. Recently, however, interest and activity have slowly revived. The Commerce Department's Office of Japanese Technical Literature—a leading player on the Task Force—recently sponsored two national conferences on Japanese-to-English machine translation. The first, held in September 1988, was designed to gauge U.S. and Japanese efforts and to determine future U.S. policy. Participants described the many Japanese commercial systems, including Bravice, CSK, Nippon Electric, Fujitsu, Hitachi,

Toshiba, Japan IBM, Matsushita, and Mitsubishi. Also highlighted were the two *primary* MT systems under development in the United States:

- A system by the Systran Corporation, under contract to the U.S. Government (i.e., the Air Force Foreign Technology Division at Wright-Patterson Air Force Base, Ohio). Considered extremely fast, Systran's system is the world's most widely used MT technology by governments and industries. Placed in service in 1969, the original system has been steadily modernized over the years. It uses the transfer mode of translation (a process outlined below), which is the most prevalent method among MT systems today.
- A "universal parser" system, an advanced multilingual system using another, newer mode of translation, known as interlingua (also described below). This system has been developed by Jaime Carbonell and Masaru Tomita at the Center for Machine Translation of Carnegie-Mellon University, with support from NSF and Japanese corporations.

[Editor's Note: *Parsing* refers to decisions about the relative functions of words in a given sentence, e.g., subject, verb, direct and indirect objects, and modifiers.]

The second conference, held in December 1989 at the National Academy of Sciences, aimed to inform the U.S. audience about Japan's advances in Japanese-to-English machine translation. Professor Makoto Nagao of Kyoto University was the keynote speaker. Nagao is an internationally recognized expert and a father of Japanese MT research, whose work has been supported by Japan's Science and Technology Agency. The symposium featured four panels, focusing on the state-of-the-art, market opportunities, users' needs, and prospects for R&D promotion. A report of the symposium is being prepared by the Academy. U.S. interest in machine translation may well have been catalyzed by the symposium.

For several years, NSF has been supporting basic research in MT (including the work of Jaime Carbonell's group, as noted above). As of FY 1990, NSF is funding three such projects, at a total level of approximately \$337,000, through the Division of Information, Robotics, and Intelligent Systems,

under the Directorate for Computer and Information Science and Engineering (CISE). These projects are:

1. Machine Tractable Dictionaries as Tools and Resources for Natural Language Processing. Principal investigator: Yorick Wilks, New Mexico State University. Award: \$131,436. Duration: 12 months.
2. Multilingual Natural Language Processing. Principal investigator: Masaru Tomita, Carnegie-Mellon University. Award: \$62,500. Duration: 12 months (Presidential Young Investigator award).
3. A Sublanguage Approach to Japanese-English Machine Translation. Principal investigator: Ralph Grishman, New York University (joint work with Michiko Kosaka, Monmouth College, and Virginia Teller, City University of New York). Award: \$142,869. Duration: 12 months.

In addition, CISE is supporting two projects in language text generation and 10 others in natural language processing.

At this time, other U.S. organizations conducting MT research include the Courant Institute of Mathematical Sciences at New York University; Martin Marietta Corporation; Microelectronics and Computer Technology Corporation; Executive Communication Systems (ECS) in Utah; IBM's T. J. Watson Research Center; Logos Corporation; and Pan American Health Organization.

U.S. potential in MT technology is judged to be strongest in linguistic analysis and the production of efficient computer code. R&D projects in which the United States and Japan can cooperate—without inordinate risk of competitive loss for either side—include analysis of word boundaries and development of input devices (particularly optical character recognition [OCR] devices) and dictionaries.

The Task Force Study on Machine Translation

To explore firsthand the possibilities inherent in M(A)T technology, the Task Force is making final preparations to send its study team to Japan, as part of the cooperation initially agreed upon by the U.S.-Japan Joint Task Force on Access at its Tokyo meeting in April 1989. The project was discussed in detail with the Japanese in early April 1990 at a Joint Task Force meeting, and several weeks later at a Joint Working Level meeting, both in Washington.

The Team of Experts. Plans call for the team to be comprised of about six members, headed by a university-based investigator, who are considered representative of the community associated with M(A)T—from fundamental research through end use. In selecting the participants, the Task Force is aiming for expertise in a wide cross-section of specialties, including linguistics, intelligent systems, users' requirements, and Japanese language. In addition, it is striving for balance among academia, the private sector, and government. The team of experts will be accompanied by one or two governmental officials.

Jaime Carbonell of Carnegie-Mellon University, mentioned earlier, has been chosen as the team's leader. His recommendations will be integral to the Task Force's final selection of the team members. As of July 1990, the full composition of the team has not yet been determined.

The Objectives. The team's immediate mission will be to:

- Assess the state-of-the-art in M(A)T. The team will evaluate JICST's Japanese-to-English system and 20 to 30 other projects, with respect to their abilities to assist American users who are unfamiliar with Japanese, their fidelity to technical intent, and the idiomatic accuracy of their output. Also of interest will be such factors as expandability to U.S.-manufactured personal computers and workstations, adaptation of OCR devices, and accessibility via electronic communications.
- Explore potential avenues of U.S. cooperation with Japanese counterparts in developing MT technology (from Japanese to English).

- Document findings and develop recommendations for future U.S. policy directions. In addition to conclusions, the Task Force is placing particular emphasis on *recommendations* as an outcome of the study.

The Task Force has also targeted several long-term goals for the U.S. Government. Strategies for attaining them will be based directly on the outcome of the study. These overall objectives are to provide leadership in:

- Identifying problems in advancing MT technology.
- Creating opportunities for U.S.-Japan cooperation in MT research.
- Increasing the flow, through machine and machine-assisted translation, of Japanese technical information to U.S. researchers.

The team's evaluation is expected to benefit several principal groups: scientists and engineers who are conducting MT research; users whose research will be expedited by access, through machine and machine-aided translation, to Japanese scientific and technical reports; translators of such reports; and developers of commercial MT systems. The special interests of these communities may overlap in some respects, and may be mutually exclusive in others. The Task Force hopes to reflect a balance of these perspectives in its selection of team participants, in addition to other aspects of diversity within the team.

Both Professor Nagao and JICST officials have expressed their willingness to assist the Task Force in selecting the centers of excellence and experts the team members should visit, and in arranging those visits. Sites will be chosen for their representation of the *best* research being carried out, regardless of whether they are affiliated with government, academia, or industry. Because there will likely be 20 to 30 research centers to visit during the short stay in Japan, the team will be divided into groups of about two per site visit; an attempt will be made to match each panelist's area of expertise with appropriate sites. However, those research centers deemed most important will be visited by the entire team.

Administration. The study will be carried out under the aegis of the Japanese Technology Evaluation Center (JTEC) Program. JTEC is managed by NSF's Division of Electrical and Communications Systems, under the Directorate for Engineering, and is administered under contract by the Japanese Technology Evaluation Center at Loyola College in Baltimore.

JTEC was initiated by NSF and the Department of Commerce in 1983 to provide definitive technical assessments of emerging Japanese trends in selected areas of high technology. These assessments are conducted by panels of five to seven senior U.S. experts from business, academia, and government. Each panel produces a report and a workshop analyzing the current status and long-term forecast of Japanese R&D in the relevant discipline. Prior JTEC panels have focused on such areas as biotechnology, advanced materials, high-definition television, superconductivity, and computer-aided design and manufacturing in the semiconductor industry. Currently, JTEC receives funding for individual studies from NSF, the National Aeronautics and Space Administration (NASA), the Departments of Commerce and Energy, and the Defense Advanced Research Projects Agency (DARPA).

JTEC participants have traditionally been well-received by their Japanese hosts, in terms of both generosity of time and openness of discussion. In selecting JTEC as the mechanism for the MT study, the Task Force considered the advantages of JTEC's reputation for establishing channels of constructive communications and exchange of ideas among Japanese and American scientists and engineers. [Editor's Note: For more information on JTEC, contact: Duane Shelton, Japanese Technology Evaluation Center, Loyola College, 4501 North Charles Street, Baltimore, Maryland 21210-2699; telephone: (301) 323-1010; FAX: (301) 532-5123. JTEC reports are available through NTIS at the address cited earlier, or by calling (703) 487-4650.]

Funding Strategy. For the study, the Task Force is considering support to JTEC at a level of about \$130,000, derived from multiple sources. This sum will be used to cover advance work, such as translation of material and meeting arrangements; transportation and travel costs; honoraria; preparation and production of the final report and

workshop; direct operational expenses, including salaries and travel costs, for JTEC staff; and indirect overhead expenses to Loyola College.

While the final funding strategy has yet to be resolved, commitments have come from the U.S. Departments of Commerce and Defense, and from NSF's CISE Directorate and Division of International Programs. Invitations to participate in the study have also been extended to private companies.

Next Steps. Once the requisite funding has been secured and the team members have been selected, the Task Force will hold a kickoff meeting to orient the group. Preparations will include a review of the research literature and confirmation of the itinerary.

The travel to Japan will constitute one week. Combined with pre- and post-visit work, roughly one month of intensive commitment will be expected of each team member. About six to eight weeks after returning from Japan, the team will conduct debriefings and seminars and assemble their results, under the guidance of the Task Force. They will draft a report, containing an executive summary, a trip report, and an analytic document integrating their observations and conclusions. As mentioned above, the Task Force has requested that the report also include recommendations about the most fruitful areas for cooperative R&D.

The draft report will then be presented at a workshop at NSF, to which the press and public (including, but not limited to, Americans and Japanese) will be invited. Modifications that result from the workshop discussion will be incorporated in a final report, which will be made available to the public through NTIS. Plans call for dissemination of the study results via an aggressive outreach to academic, corporate, and governmental research organizations. NSF is investigating the potential use of teleconferencing and videotape media. The entire project must be completed by March 1991; in April, the next round of U.S. Task Force meetings with Japanese counterparts will take place in Tokyo.

More About Machine Translation Technology

While machine translation is a highly complicated, interdisciplinary area of research, non-experts can readily gain an appreciation for its sophistication and for its enormous potential benefits for global society.

An Imperfect Solution. Machine translation is not a straightforward matter of substituting words and syntax from one language to another. Take, for example, the sentence: "I ordered hot tea and pastries." A computer system might produce the sentence: "I ordered hot tea and hot pastries." While this may not be an egregious misinterpretation, in many cases the result can be garbled indeed.

From Japanese to English, human intervention at both ends of the process is required with most current MT applications. Pre-editing may be labor-intensive, but it has a demonstrably important role when the material is going from Japanese (which can have a random word order and infrequent boundaries between words) to English (in which the word order is more structured and words are discrete). Post-editing will further clarify ambiguities and create a more polished text.

The Translation Process. What characterizes a Japanese-to-English MT system? As mentioned earlier, there are two principal modes in use: the transfer method, which is currently the more popular, and the method that uses an interlingua. With the *transfer* method, in addition to the colloquial and technical dictionaries cited earlier, the salient features of most systems are:

- Input of a sentence after pre-editing.
- Morphological analysis of a string of Japanese characters to form meaningful words.
- Syntactic analysis of these words to form a syntactic dependency structure.
- Semantic analysis of the syntactic dependency structure to form a semantic dependency structure for Japanese.
- Transformation of the Japanese semantic dependency structure into one of English.
- Syntactic generation of an English phrase (or tree) structure.

- Morphological synthesis of an English sentence with appropriate tense, articles, word position, number, etc. (This function is complicated because Japanese lacks articles and is often imprecise in plurality.)
- Output of a translated English sentence, followed by post-editing.

Besides having this basic construct, most MT programs also include algorithms for broader capabilities, such as contextual analysis or intersentential relationships.

A system that uses an *interlingua* will have most, if not all, of the above components. The crucial difference with this approach is that the output of the syntactic and semantic analysis is translated into a "pivot" language. Any pair of languages must transit through this pivot, rather than being converted directly from one to the other. The benefit of the interlingua process is that the number of programmed dictionaries can be greatly reduced, because the interlingua code acts as the hub of a wheel and individual languages as the spokes radiating from it. For example, for text to be translated from Japanese into French, the computer converts it into interlingua first and then into French. Thus, while programmers must create dictionaries and syntactical constructs for the Japanese-interlingua and French-interlingua language pairs, separate dictionaries and syntactical constructs are unnecessary for Japanese-French, Japanese-German, or Japanese-Spanish. The Japanese-to-English system developed at Carnegie-Mellon is based on the interlingua approach.

Linguistic Challenges for Intelligent Systems. The complexities of language are sufficient to fill many analytical journals. An idea of the challenges they pose for a machine translation system, however, can be suggested by a few examples. Written Japanese consists of two phonetic syllabaries called *kana*, each with 46 characters. In addition, there is a third script, called *kanji*, which consists of Chinese characters. At least 10,000 *kanji* characters exist, of which about 3,000 are commonly used in lay publications (e.g., newspapers); another 1,000 to 2,000 serve the specialized vocabularies of science, technology, and medicine. These orthographies require vastly more computer storage space than do the 26 characters of the English alphabet.

Linguistic ambiguities pose difficult questions and place tremendous demands on systems to "think." Homonyms—words with identical spellings but different pronunciations and meanings—are especially tricky (for example, the English words "read" and "lead" may be variously interpreted). Similar challenges are presented by homographs, or words with more than one part of speech: "flies," for instance, can be either a plural noun or a singular verb. And polysemes, which are words with more than one meaning for the same part of speech, can compound a machine's potential confusion: "clutch" can refer alternately to a mechanical device for engaging an automobile's transmission, or to the eggs produced by a bird or reptile—no small disparity.

One of the goals being sought in MT research is for a Japanese-to-English system to be able to sort out the conflicting possibilities and try to match the translation with the proper meaning, based on a program of weighted word relationships stored in a simulated neural network. The "weights" assigned to possible meanings are based on the context in which the original word appears; from them, the computer makes probable associations. A senior research scientist at Toshiba's Information Systems Laboratory, cited in *New Scientist*, used an example from a Japanese *kanji* text which contains the sentence:

The President is to enact the bill as soon as it passes through the legislature.

The next sentence of the Japanese text begins with the word *shiko*, which has many definitions, including "trial," "enforcement," "orientation," and "consideration." The system makes an informed guess and selects the correct word, "enforcement," because it has the strongest bonds with the words "enact" and "bill" in the first sentence.

Similarly, in the paradigm that precedes this article, the French word *avocat* is a polyseme, meaning either "avocado" or "advocate." A translation machine whose dictionary has been programmed to make weighted linkages by subject matter would select "advocate" as the meaning, because of the probable association with legal terminology suggested by "court." As Peter J. Wheeler, the expert who actually encountered the "avocado" translation a dozen years ago, remarked in "The Errant Avocado," an article published in the British

Computer Society's *Newsletter of the Natural Language Translations Specialist Group* in February 1983: "Suddenly that intriguing fruit, the general avocado, has been supplanted by the reassuringly solemn figure of an advocate-general. We can all breathe again, and go back to enjoying our avocados and respecting our advocates." [Editor's Note: Wheeler is president of Antler Translation Services of Sparta, New Jersey, one of the few translation companies actually using MT.]

Conclusion

While the advent of a fully automatic system may be many years away, the attraction of MT technology lies in its promise to reduce an otherwise unfathomable burden for human translators in dealing with massive technical material. With advances in knowledge, the proverbial flaws in machine translation which today may evoke our amusement—as with the anecdote about the general avocado—might one day be a thing of the past. Such progress will surely be enhanced by future U.S.-Japanese cooperation in this field. Bringing the two countries together to advance MT research would be a major and lasting contribution of the Task Force study team.

Sources: Participation as a member of the U.S.-Japan Task Force on Scientific and Technical Information; Public Law 99-382 of August 14, 1986; Executive Order 12591 of April 10, 1987, "Facilitating Access to Science and Technology"; *Japanese Technical Literature Bulletin*, November 1988, January 1989, August 1989, Office of Japanese Technical Literature, U.S.

Department of Commerce; Task Force Draft Report of April 6, 1990; documents from the Japanese Technology Evaluation Center, Loyola College, Maryland; "Report to Congress: Current Status of Japanese-to-English Machine Translation," Tamami Kusuda et al., U.S. Department of Commerce, July 1988; testimony of Christopher T. Hill, Congressional Research Service, and Tomoyuki Satoh, T. Satoh & Company, before the U.S. House of Representatives Subcommittee on International Scientific Cooperation, Committee on Science, Space, and Technology, April 27, 1988; Christopher T. Hill, "Japanese Technical Information: Opportunities to Improve U.S. Access," Congressional Research Service Report for Congress #87-818 S, October 13, 1987; *Language and Machines: Computers in Translation and Linguistics*, report by the Automatic Language Processing Advisory Committee (ALPAC), Washington, D.C.: National Academy of Sciences/National Research Council, 1966; Hirotsato Nomura et al., *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, U.S.A.*, Tokyo: Japan Electronic Industry Development Association, July 1989; Tokyo trip report by Joseph E. Clark, National Technical Information Service, April 24-28, 1989; discussions with Muriel Vasconcellos, Pan American Health Organization, May and June 1990; personal communication with Peter J. Wheeler, Antler Translation Services, June 1990; personal communication with Mark F. Eaton, Microelectronics and Computer Technology Corporation, March 1988; *Summary of Awards, Fiscal Year 1989*, NSF Division of Information, Robotics, and Intelligent Systems, Directorate for Computer and Information Science and Engineering; *Science & Engineering Indicators—1989*, National Science Board, Washington, D.C., 1989; Foreign Broadcast Information Service report; *New Scientist*, January 6, 1990; *TIME*, July 24, 1989.

Technology

Trying to Decipher Babel

Japanese translating machines make languages less foreign

BY BARRY HILLENBRAND TOKYO

The machine clearly does not like poetry. It won't touch the stuff. Nor is it very fond of novels. Theoretically, it could cope with some of Hemingway's short, simple sentences, though it could never make anything of long, convoluted passages from Faulkner. But give the Toshiba AS-TRANSAC computer a thoroughly dull, straightforward instruction manual, and it will earnestly chomp its way through page after page. What it does with those pages is the amazing part. The Toshiba machine has linguistic ability far beyond the powers of past generations of computers: it can translate, at least crudely, one language into another. In this case, the computer converts simple English into serviceable, if stilted Japanese.

The AS-TRANSAC is one of more than half a dozen machine-translation systems being energetically developed in Japan. With their strong thirst for information from other nations and a growing need to disseminate their documents around the world, the Japanese urgently require computers that can translate. A few machines, such as the Toshiba model and Fujitsu's Atlas system, are already in operation, helping Japanese companies like Mazda translate technical material. A powerful computer called SHALT, designed by IBM Japan, is being used extensively for in-house translations. In 1988 SHALT converted four IBM manuals from English into Japanese. This year the target is 20 to 30. Predicts Kiyotaka Yasui, manager of the language and image-technology section at IBM Japan's Tokyo research laboratory: "In five years the internal-publication department of IBM Japan will be fulfilling 100% of its translation requirements via machines."

But human translators should have no fear that their jobs are imperiled—at least for now. None of the new systems are yet able to take a page of text and render it unerringly into a different language without the aid of a bilingual editor who can fine-tune the output for ambiguities in the vocabulary, to say nothing of shades of meaning. "A truly automatic system is a dream at the moment," admits Makoto Ihara, manager of Toshiba's computer product-planning department. Says Kazunori Muraki, a leading researcher at NEC: "Machine translation is only to reduce the work involved in human translation."

And that it does. The present generation of machine-translation systems,

which are priced between \$30,000 and \$70,000, can nearly double the output of translators of technical documents. The savings, especially for small firms unable to maintain a large staff of skilled translators, can be considerable.

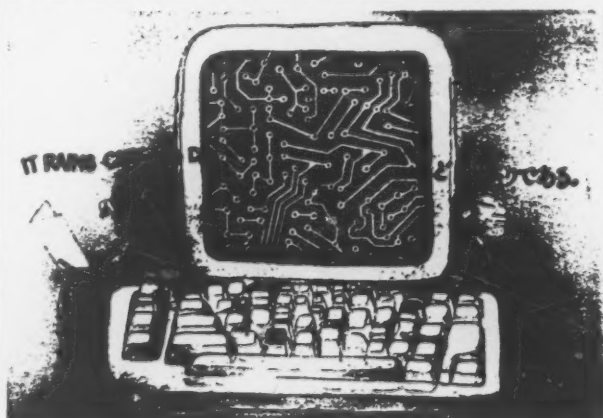
"Seven or eight years ago," says Koichi Takeda, a researcher at IBM Japan, "everyone was saying machine translation was a technology of the future. But now we have it."

Considering the complexity of the task, the progress in machine translation has been startling. Essentially, the trans-

Japanese student of English. The articles and subjects are gone, and the verb dangles clumsily at the end. Only after the English sentence has been transformed into Japanese syntax are the words translated.

Japan's computer makers are developing machines that can translate freely among several different languages. Fujitsu, for example, has a prototype called Atlas 2 that can deal with Japanese, French, German and English. In the near future, Spanish, Chinese and Korean will be added. To make such systems as simple as possible, programmers have invented a coded, largely numerical language called "interlingual."

Instead of translating directly from Japanese into German, the computer would translate from Japanese into inter-



lating machine analyzes the syntax of an English sentence, determining its grammatical structure and identifying, for example, the subject, verb, objects and modifiers. Then the words are translated by an English-Japanese dictionary. Next, another part of the computer program analyzes the resulting awkward jumble of words and meanings, and generates an intelligible sentence based on the rules of Japanese syntax and the machine's understanding of what the original English sentence meant.

That is not as simple as it sounds (assuming it sounds simple at all). Each computer company has devised strikingly different sets of programs to deal with the fiendish complexities of the two languages. One step in the IBM system, for example, refashions the English sentence structure and word order to resemble Japanese syntax. The result occasionally reads like the faulty work of a ninth-grade

lingual and then into German. This process cuts down on the number of dictionaries that programmers have to construct. A Japanese-interlingual dictionary would be needed, but not a Japanese-German, Japanese-French or Japanese-Spanish. Explains Hiroshi Uchida, a researcher at Fujitsu: "If we did not use interlingual, then each pair of languages would require the development of a specific set of grammatical rules and a bilingual dictionary. Interlingual acts as the hub of a wheel."

The market for such machines will be vast. Says Yasuyo Kikuta, a researcher in artificial intelligence at Fujitsu: "Since we Japanese have so much trouble in the area of foreign languages, machine translation is the kind of tool all Japanese desire." And since many people in other nations are not linguistic whizzes either, sales of the electronic translators should be brisk around the world.